ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

IDENTIFICATION OF NOVEL INTERACTION PARTNERS OF NUCLEAR FACTOR I B

M.Sc. THESIS

Ahmet Erdal Yılmaz

Department of Advanced Technologies

Molecular Biology, Biotechnology and Genetics Programme

JUNE 2013

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

IDENTIFICATION OF NOVEL INTERACTION PARTNERS OF NUCLEAR FACTOR I B

M.Sc. THESIS

Ahmet Erdal YILMAZ
(521091104)

Department of Advanced Technologies

Molecular Biology, Biotechnology and Genetics Programme

Thesis Advisor: Assist. Prof. Dr. Aslı KUMBASAR

JUNE 2013

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

NÜKLEER FAKTÖR I B'NİN ETKİLEŞİM PARTNERLERİNİN TANIMLANMASI

YÜKSEK LİSANS TEZİ

Ahmet Erdal YILMAZ
(521091104)

İleri Teknolojiler Anabilim Dalı

Moleküler Biyoloji-Genetik ve Biyoteknoloji Yüksek Lisans Programı

Tez Danışmanı: Yrd. Doç. Dr. Aslı KUMBASAR

HAZİRAN 2013

Erdal Yılmaz, a **M.Sc.** student of ITU **Graduate School of Science, Engineering and Technology,** student ID **521091104** successfully defended the **thesis** entitled "**IDENTIFICATION OF NOVEL INTERACTION PARTNERS OF NUCLEAR FACTOR I B"** which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assist. Prof. Dr. Aslı KUMBASAR** Istanbul Technical University	.....
Jury Members :	**Assist. Prof. Dr. Gizem Dinler Doğanay** Istanbul Technical University	.....
	Assist. Prof. Dr. Şenay Vural Korkut Yıldız Technical University	.....

Date of Submission : 03 May 2013
Date of Defense : 07 June 2013

To all scientists and my friends,

FOREWORD

I specially thank to Ceren ALKIM and Ülkü YILMAZ for sharing their knowledge base with me about yeast, to Şeyma KATRİNLİ and Umut GÜNSEL for allowing me to use their equipment, to Nesteren MANSUR for her laboratory friendship, to Koray KIRIMTAY and Ece SELÇUK for sharing their scientific knowledge with me, to Aydın ÖZMALDAR for his advices in thesis writing, to Meryem MENEKŞE and Havva Esra BIYIK for allowing me to use their equipment, to Bilal TETİK for providing biological and other kinds of materials.

Lastly, I am thankful to ITU Research Funds for economically supporting this thesis project.

June 2013 Ahmet Erdal YILMAZ

TABLE OF CONTENTS

Page

ABBREVIATIONS

AD : Activation Domain
Amp : Ampicillin
cDNA : Complementary DNA
DDO : Double drop-out medium
DDO/X : Double drop-out medium with X-α-gal
DDO/X/A : Double drop-out medium with X-α-gal and aureobasidin A
DMSO : Dimethyl Sulfoxide
DNA-BD : DNA Binding Domain
EDTA : Ethylene Diamine Tetraacetic acid
EtBr : Ethidium Bromide
GAL4 : Galactose promoter binding transcription factor
His : Histidine
Kan : Kanamycin
LB : Luria Bertani
Leu : Leucine
LiAc : Lithium Acetate
mRNA : Messenger ribonucleic acid
NCBI : National Center for Biotechnology Information
PCR : Polymerase Chain Reaction
PEG : Polyethylene glycol
SD : Synthetic Dextrose
SDS : Sodium dodecyl sulfate
TAE : Tris-acetate EDTA
Taq : Thermus aquaticus
TE : Tris-EDTA
Trp : Tryptophan
QDO : Quadruple dropout medium
QDO/A : Quadruple dropout medium with aureobasidin A
QDO/X/A : Quadruple dropout medium with X-α-gal and aureobasidin A
X-α-gal : 5-Bromo-4-chloro-3-indoxyl-α-D-Galactopyranoside
YPD : Yeast Peptone Dextrose

LIST OF TABLES

Page

LIST OF FIGURES

Page

IDENTIFICATION OF NOVEL INTERACTION PARTNERS OF NUCLEAR FACTOR I B

SUMMARY

Nuclear factor I (NFI) family of transcriptional factors has four members (Nfia, Nfib, Nfic, and Nfix) and is a highly conserved gene family, which plays a key role in central nervous system development. These proteins are required for glial and neuronal cell differentiation and proper development of neuronal and glial populations in developing brain. By post-transcriptional processing many different splice variants can be produced from a single NFI gene and this creates additional heterogenity in NFI proteins.

The typical NFI protein is composed of an N-terminal DNA-binding/dimerization domain and a C-terminal transcriptional activation and/or repression domain. The N-terminal domain is around 200 amino acids and responsible for DNA-binding activity, dimerization and the stimulation of adenovirus DNA replication. The C-terminal domain is referred to as a proline-rich activation domain and it is mostly responsible for regulation of gene expressions. NFI proteins bind to duplex DNA with TTGGC(N_5)GCCAA palindromic motif. NFI proteins form either heterodimers or homodimers and bind DNA with similar affinity.

During embryogenesis, NFI genes are expressed in distinct, overlapping patterns with high levels of expression of Nfia, Nfib, and Nfix in the developing central nervous system in ventricular zones. At late embryonic and early postnatal ages, expression is particularly high in the cerebral cortex, hippocampus, basilar pontine nuclei, cerebellum, and spinal cord. While NFI-A and NFI-B are expressed during early gliogenesis, NFI-X and NFI-C are expressed later during differentiation of astrocytes and control the expression of late astrocyte markers. NFI proteins also regulate the cell migration and terminal differentiation in the cerebellum and glia mediated cell migration in the hippocampus. NFI family members are also involved in neurological diseases and cancer.

The yeast two-hybrid system is a genetic system wherein the interaction between two proteins of interest is detected via the reconstitution of a transcription factor and the subsequent activation of reporter genes under the control of this transcription factor. Since interacting partners of NFIB in developing brain have not been well studied we aimed at identifying NFIB binding proteins by screening a human fetal brain cDNA library with the yeast two hybrid technique. As a result, we identified 12 potential NFIB interacting proteins of NFIB in fetal brain library.

NÜKLEER FAKTÖR I B'NİN ETKİLEŞİM PARTNERLERİNİN TANIMLANMASI

ÖZET

Nükleer faktör I (NFI) transkripsiyon faktörleri ailesi barındırdığı 4 üyesi ile birlikte yüksek oranda korunumlu ve merkezi sinir sistemi gelişiminde anahtar rol oynayan bir ailedir. Bir diğer ismi de CTF ya da CCAAT-bağlanma transkripsiyon faktörleri olan NFI ailesinin, tek hücreli organizmaların genomunda yer almamasından dolayı metazoalara spesifik olduğu söylenebilir.

NFI ailesi temelde 4 gen (Nfia, Nfib, Nfic, ve Nfix) ile temsil edilmesine rağmen tek bir genden post-transkripsiyonel modifikasyon ile 9 adet ürün üretilebilmektedir. NFI ailesi üyelerinden üretilen tüm proteinler N-uçlu DNA-bağlanma/dimerleşme domeyni ve C-uçlu transkripsiyonel aktivasyon ya da represyon domeyni olmak üzere iki temel domain içermektedir.

N-uçlu DNA-bağlanma/dimerleşme domeyni 200 aa uzunluğundadır. Yapısal olarak kendine hastır ve yalnızca SMAD proteinlerinin MH1 (Mad Homoloji) domeyni ile sekans homolojisi taşımaktadır. İnsan, fare ve tavuk NFI genleri arasında yüksek oranda korunmuş olmakla birlikte DNA'ya bağlanma, dimerleşme ve adenovirüs DNA'sını replikasyonu stimüle etme gibi görevleri mevcuttur.

C-uçlu domeyn, yaklaşık yüzde yirmibeşinin prolin rezidülerinden ibaret olmasından ötürü prolince zengin domeyn olarak adlandırılır ve çoğunlukla NFI aracılı gen ekspresyonlarından sorumludur. N-uçlu domeynin tersine, C-uçlu domeyn NFI proteinleri arasında belirgin farklılıklar göstermektedir ve bu farklılıkların C-uçlu domeynin transkripsiyonel fonksiyonlarını belirlediği düşünülmektedir.

Transfeksiyon çalışmalarıyla NFI-A ve NFI-X C-uçlu domeynlerinin transkripsiyonu baskıladığı görülmüştür. Ayrıca NFI proteinlerinin pek çok promotoru negatif olarak regüle ettiği görülmüştür. NFI proteinlerinin bu şekilde hem transkripsiyonu baskılama hem de aktifleştirme rolünün olmasından ötürü, bu proteinlerin aktifleştirme/baskılama mekanizmasının eksprese edildikleri hücre tipine ve bağlandıkları promotora bağlı olduğu düşünülmektedir.

NFI proteinleri çift iplik DNA üzerinde TTGGC(N_5)GCCAA dizisine bağlanmakta ve NFI proteinlerinin insan genomik DNA'sı üzerinde toplamda 75.000 bölgeye bağlandığı tahmin edilmektedir.

NFI transkripsiyonel faktörleri ailesinin ilk olarak Adenovirus DNA'sının replikasyonundan sorumlu olduğu ve fare memeli tümör virüsünün transkripsiyonunda rol aldığı gösterilmiş, daha sonraları ise pek çok hücresel genin transkripsiyonunda ve ekspresyonunda görev aldığı anlaşılmıştır. NFI proteinlerinin ayrıca histon H1 ve H3 ile de etkileştiği gösterilmiştir.

NFI mRNA'larının vücuttaki pek çok organda örtüşen şablonlarla eksprese edilmeleri bu proteinlerin fonksiyonel analizlerini karmaşık hale getirmektedir.

NFI ailesi üyelerinden Nfib, akciğer epitelyumundaki genlerin ekspresyonu regüle etmesinden ötürü akciğer gelişiminde özel bir öneme sahiptir. NFI ailesinin bir diğer üyesi olan Nfia kalp, karaciğer, akciğer, yumurtalık, iskelet kası, böbrek, testis, pankreas, dalak, fetal karaciğer ve beyin gibi pek çok organda eksprese edilmektedir. Nfic ise özellikle diş kökü oluşumunda rol almaktadır.

Gelişmekte olan beyinde önce nörogenez daha sonra gliogenez meydana gelmektedir. Nöral gelişimin bu sıralı yapısı dışsal sinyaller, farklı transkripsiyon ailesi faktörleri ve histon ile DNA'nın modifikasyonu tarafından kontrol edilmektedir. NFI transkripsiyon faktörleri ailesi de merkezi sinir sistemi gelişiminde glial ve nöronal hücre faklılaşmasına rehberlik eden önemli bir aktör konumundadır.

Merkezi sinir sisteminin gelişimi sırasında meydana gelen nöron-glia geçişi tam anlaşılmış bir proses olmamakla birlikte Notch sinyalizasyon yolağının temel bir rol üstlendiği bilinmektedir. Ancak Notch yolağının aktif hale gelmesi nöral progenitör hücrelerinde Nfia ekspresyonuna bağlıdır. Nfia'nın bir diğer rolü de gliogenezin koordinasyonu için Sox9 proteini ile birlikte nörogenezi sabitlemesidir. *Nfia,* gliogenez ve oligodentrosit faklılaşmasını modüle etme görevini sadece beyinde değil gelişmekte olan omurilikte de yürütmektedir. *Nfib*$^{-/-}$ farede yapılan çalışmalar gelişmekte olan neokorteks ve hipokampüste glial gelişimde gecikme olduğunu göstermiştir. Bu da beyindeki glial gelişim sırasında Nfib'nin de görev aldığını ortaya koymaktadır.

Anatomik açıdan bakıldığında, Nfia, Nfib, ve Nfix başlangıçta ventriküler bölgelerde ifade edilirken, geç embriyonik ve postnatal dönemlerde ise serebral korteks, hipokampus, beyincik, ve omurilik da ekspresyon yüksek olmaktadır. NFI proteinleri ayrıca beynin gelişim sürecindeki hücre göçünü de düzenlemektedirler.

NFI ailesi üyeleri ayrıca kanser ve nörolojik hastalıklarla da ilişkilidir.

NFI proteinlerinin tavuk embriyosu fibroblast hücrelerinde aşırı ekspresyonu yapıldığında fibroblast hücrelerini onkogenlerin transformasyonuna daha dirençli hale getirdiği görülmüştür. Spesifik olarak *NFI-A* geninin farklı kronik kötü huylu miyeloid hastalıklarıyla ilişkili olduğu ve *NFIB*'nin Adenoid kistik kanserindeki rolü gösterilmiştir.

Meme kanserli dokularda NFIB proteininin ekspresyon sıklığı çalışılmış ve NFIB mRNA düzeyinin triple negatif meme kanserinde belirgin miktarda olduğu gözlenmiştir. Diğer taraftan NFIB geninin susturulmasının hücre çoğalmasında azalma ve apoptotik sinyal yolağında artışa yol açtığı gösterilmiştir.

NFI ailesinin nörolojik hastalıklardaki rolüne bakıldığında *Nfia* geninin homozigot delesyonunun hayvanlarda %95 ölüme yol açtığı ve sağ kalanların da ciddi nörolojik kusurları olduğu gösterilmiştir. Yine *Nfib*$^{-/-}$ farede beyincik içerisinde gelişim defektleri tespit edilmiştir. Nfix geninin susturulduğu farelerde ise perinatal ölüm hidrosefalus, korpus kallosumun oluşum bozukluğu, beyin kütlesinde artış ve anormal hipokampüs oluşumu rapor edilmiştir.

Bu tez çalışmasında Maya 2 Hibrid (M2H) tekniği kullanılarak insan fetal beyin kütüphanesinde Nfib3 proteininin bağlanma partnerleri araştırılmıştır. M2H sisteminde iki protein arasındaki etkileşim, bir transkripsiyon faktörünün yeniden inşa edilmesi ve devamında bu transkripsiyon faktörünün kontrolü altındaki haberci genlerin aktivasyonu ile tespit edilir. NFIB ile etkileşecek proteinlerin tespit edilesinde M2H tekniğinin seçilmiş olmasında tekniğin görece ucuz ve basit olması,

aynı zamanda maya sistemlerinin yüksek ökaryotlarda rastlanan posttranslasyonel modifikasyonlara imkan tanıması rol oynamıştır.

Bu tez projesinde "Matchmaker" M2H sistemi kullanılmıştır. Bu sistemde AUR1-C, HIS3, ADE2, MEL1 olmak üzere 4 adet haberci gen bulunmakta ve protein etkileşimlerini tespit etmek için sırasıyla Aureobasidin A antibiyotiğine dirençlilik, histidin amino asiti yoksunluğunun giderilmesi, adenin amino asiti yoksunluğunun giderilmesi ve X-α-Gal substratının parçalanarak mavi renk açığa çıkartılması esaslarından faydalanılmıştır.

Bu projede öncelikle Nfib3 cDNA'sı ve Nfib3'ün kısaltılmış versiyonu olan NFIBΔ, pGBKT7 ekspresyon plazmidine klonlanmıştır. Sonrasında Nfib proteininin eksprese edildiği, Nfib fragmentlerinin mayaya toksik etki göstermediği ve ilgili haberci genleri otoaktive etmediği gibi bir dizi kontrol deneyi yapılmıştır. Bu kontrol deneyleri sırasında kısaltılmış versiyonun otoaktivasyon gösterdiği tespit edilmiş ve esas maya 2 hibrit taramalarına yalnızca Nfib3 tam fragmenti ile devam edilmiştir.

Maya 2 hibrit taramaları için insan fetal cDNA kütüphanesi taşıyan 'av' maya suşu, Nfib3 cDNA'sı taşıyan 'yem' maya suşu ile 1 gün boyunca eşleştirilmiş ve elde edilen diploid mayalar petrilere ekilerek ilk seleksiyon AUR1-C ve MEL1 haberci genleri üzerinden gerçekleştirilmiştir. Sonrasında HIS3 ve ADE2 haberci genlerinin de katıldığı seleksiyon ve bunu takiben maya koloni PCR, restriksiyon kesimi, ve segregasyon analizi ile 'gerçek etkileşim'lerin ortaya çıkartılması hedeflenmiştir. Son olarak, bulunan ekileşimlerin teyit edilmesi amacıyla bir eşleştirme çalışması daha yapılmıştır. Bu son deney ile elde edilen klonlar dizi analizine gönderilerek NFIB proteini ile etkileşen proteinlerin cDNA'sı tespit edilmiş ve veritabanlarıyla karşılaştırılarak bunların hangi genlere denk geldiği ortaya çıkarılmıştır.

INTRODUCTION

1.1. Background Information

1.1.1. NFI family of transcription factors

The nuclear factor I (NFI) family of transcription factors is composed of four members, Nfia, Nfib, Nfic, and Nfix which are highly conserved among vertebrates. These transcription factors are specific to metazoans as no members of this family is present in the genome of a unicellular organism. NFI family members have also been referred to as the CTF or CCAAT-binding transcription factors.

Up to now, NFI proteins have been found to have binding sites on and to regulate many genes (Gronostajski et al., 2011) and NFI proteins are estimated to bind to ≈75.000 sites in total on human genomic DNA (Gronostajski et al., 1985). DNA-binding assays (Hennighausen et al., 1985; Leegwater et al., 1985; Nowock et al., 1985) and isolation of NFI binding sites (Gronostajski et al., 1985) revealed that NFI proteins bind to a palindromic motif with five or six nucleotides spacing between two pentameric half-sites (TTGGC(N_5)GCCAA sequence) on duplex DNA. However it was also shown (Gounari et al., 1990; Wuarin et al., 1990) that individual half-sites (TTGGC or GCCAA) are sufficient for promoter binding and transcriptional activation. The binding affinity of NFI proteins for half sites ($K_d \sim 10^{-9}$ M) is remarkably reduced compared to that of the dyad symmetric site ($K_d \sim 10^{-11}$ M) (Meisterernst et al., 1988). NFI proteins form either heterodimers or homodimers (Kruse and Sippel, 1994b; Chaudhry et al., 1998). The binding affinity of heterodimers and homodimers of NFI proteins to DNA were investigated for all four NFI genes and no significant difference was detected (Kruse and Sippel, 1994b). It was also shown that transcriptional activation potentials of NFI heterodimers are at intermediate level when compared with NFI homodimers and it was proposed that this could be a potential mechanism by which different NFI proteins can regulate gene expression (Chaudhry et al., 1998).

1.1.2 Structural properties of NFI family members

While NFI proteins are encoded by four genes, as many as 9 different splice variants are produced from a single gene by post-transcriptional processing and this creates additional heterogeneity in NFI proteins (Altmann et al., 1994; Chaudhry et al., 1997; Kruse et al., 1991; Kruse and Sippel, 1994a; Santoro et al., 1988; Paonessa et al., 1988).

The proteins expressed from all members of NFI family contain two distinct components: (1) an N-terminal DNA-binding/dimerization domain and (2) a C-terminal transcriptional activation and/or repression domain.

The N-terminal domain is responsible for DNA-binding activity, dimerization and the stimulation of adenovirus DNA replication (Mermod et al., 1989; Gounari et al., 1990). Deletion studies (Mermod et al., 1989; Gounari et al., 1990) showed that this DNA binding/dimerization domain is around 200 amino acids in length and is highly conserved between human, mouse and chicken NFI genes.

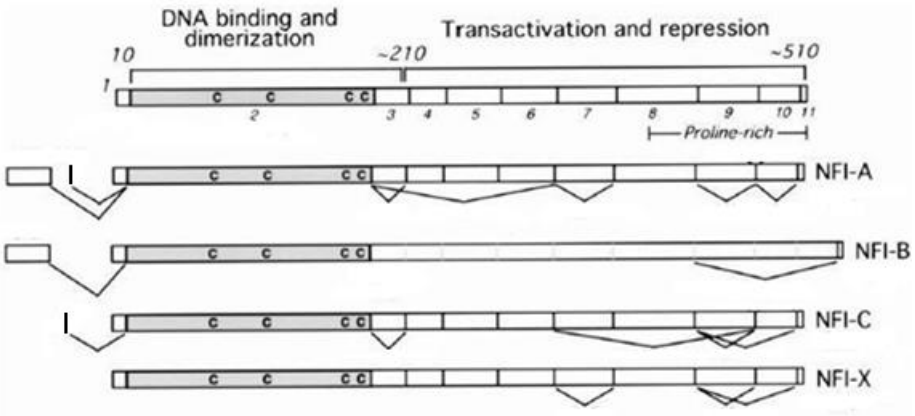


Figure 1.1 Domains and alternative splicing of vertebrate NFI genes. The top line shows general structure of NFI genes. The angle brackets below each gene represent alternatively spliced isoforms of individual family members. The boxes or lines connected to the second exon show alternative first exons. The second exon includes four conserved cysteine residues (labeled C) which are required for DNA-binding and redox regulation of binding. Approximate residue numbers and exon numbers are shown above and below the line, respectively. The C-terminal regions of each protein correspond to specialized domains (labeled Transactivation and repression) including the proline-rich transactivation domain (labeled Proline-rich) (Adapted from Gronostajski, 2000).

NFI DNA-binding domain has sequence homology to MH1 (Mad Homology) domain of SMAD proteins which implies that there may be a structural similarity as well (Sadreyev and Grishin, 2003). No other sequence homology has been reported

and NFI protein structure has not been elucidated. It was shown (Bandyopadhyay and Gronostajski, 1994) that the four cysteine residues of NFI DNA-binding domain are conserved between NFI family members and 3 of these 4 residues are needed for providing DNA-binding activity.

The C-terminal domain contains a proline-rich activation domain as approximately 25 percent of this domain consists of proline residues (Figure 1.1, Proline rich). In contrast to the DNA-binding domain, significant variations are seen within the C-terminal transactivation domains among NFI proteins and this variation of C-terminal domain structure affects transcriptional modulatory functions of NFI proteins (Santoro et al., 1988; Kruse and Sippel, 1994b). The 100 residues on the C-terminal domain corresponding to residues between 399-499, were shown to be necessary for maximal transcriptional activation of an NFI-site containing promoter in *Drosophila* Schneider cells (Mermod et al., 1989).

While the N-terminal domain is responsible for DNA-binding and replication activities of NFI proteins, C-terminal domain is mostly responsible for regulation of gene expression. With transient or stable transfection assays, the C-terminal domain of both NFI-A and NFI-X has also been shown to repress transcription (Nebl and Cato, 1995; Monaci et al., 1995; Osada et al., 1997a,b). Additionally, NFI binding sites have been shown to act as negative regulatory elements in several promoters (Adams et al., 1995; Macleod and Plumb, 1991; Osada, 1997a; Rajas et al., 1998; Gao et al., 1996; Szabo et al., 1995; Cooke and Lane, 1999).

1.1.3 Functional properties of NFI family members

The NFI family of transcriptional factors plays extensive roles in animal development and pathology. NFI protein was first shown to be responsible for the replication of Adenovirus DNA (Nagata et al., 1982) and was also reported to have an important role in transcription of mouse mammary tumor virus (Cato et al., 1988). Later it was understood that NFI family is involved in the transcription and expression of cellular genes as well (Jones et al.,1987; Jackson et al., 1993; Grewal et al., 1992). NFI proteins may activate or repress transcription depending on the specific isoform present in a particular cell type regulating different promoters, with their binding partners (Gronostajski, 2000).

The four NFI genes were shown to be expressed in overlapping patterns in several organs throughout the body complicating functional analysis of their protein products (Chaudhry et al., 1997). While all NFI family members are expressed in the lung (Chaudhry et al., 1997), Nfib has a particular importance in lung development (Mason et al., 2009; Steele-Perkins et al., 2005) and can directly regulate expression of epithelial cell specific genes in the lung (Bachurski et al., 2003). Phenotypes exhibited by *Nfic* $^{-/-}$ mice (Steele-Perkins et al., 2003; Park et al.,2007) indicate that *Nfic* is particularly important in tooth root formation. Nfic may also have a role in the regeneration and tissue growth processes of skin, hair follicles, and teeth (Plasari et al., 2010). While Nfic regulates hair growth cycle, Nfib was recently found to control hair follicle stem cell behavior (Chang et al, 2013).

A dual reporter gene assay has revealed that Nfic can protect a gene from silencing when its DNA-binding sites are inserted between the gene and the telomeric ends (Esnault et al., 2009). Ferrari et al. (2004) reported that Nfic may bind histone H3 and act as a barrier between telomeres and downstream gene.

The role of NFI proteins in adipogenesis has been shown as NFI-binding sites have been identified in an adipocyte-specific enhancer (Graves et al., 1991) and an adipocyte-specific promoter (Singh and Ntambi, 1998). Also, NFIA and NFIB were shown (Waki et al., 2011) to have physiological roles in adipocyte differentiation as siRNA knockdown of either NFIA or NFIB led to the suppression of induction of the expression of the adipogenic transcription factors PPARc and C/EBPa and also to reduction of lipid accumulation in adipogenic cell lines.

1.1.4 Roles of NFI family in CNS development

In vertebrates, the central nervous system (CNS) is consists of two main cell types: neurons and glia. The glial cells are further divided into subcategories as oligodendrocytes, astrocytes and microglia.

While the main functions of oligodendrocytes in CNS are forming myelin sheaths that insulate axons and promoting the saltatory conduction of electric signals, astrocytes are responsible for structural integrity of the brain, providing metabolic support to neurons, maintaining water and ionic balance and modulating synaptic transmission (Guillemot, 2007b).

The numerous subtypes of neurons as well as oligodendrocytes and astrocytes arise from multipotent progenitors, a single layer of neural stem cells in the early embryo, located in the walls of the embryonic neural tube. (reviewed in Guillemot, 2007b and Corbin et al., 2008). This is achieved by first neurogenesis followed by gliogenesis (Miller and Gauthier, 2007). These sequential phases of neural development are controlled by extrinsic signals, different families of transcription factors and modifications of histones and DNA (reviewed by Rowitch, 2004; Guillemot, 2007a).

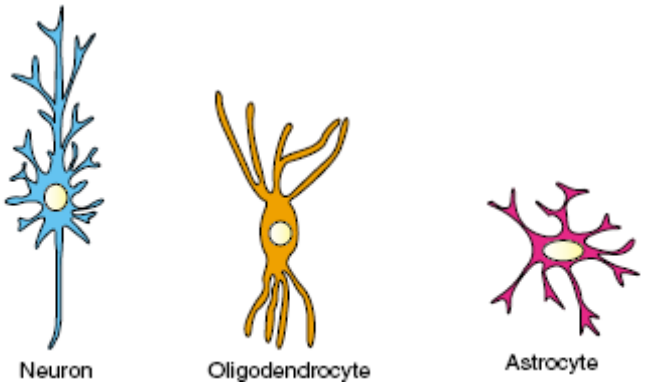


Figure 1.2 Cell types and differentiation in CNS (Adapted from Guillemot, 2007b)

The generation of a particular class of neuron or glial cell from a multipotent progenitor consists of sequential steps (Guillemot, 2007b). The whole process starts with the spatial patterning of the neural primordium in which progenitor cells acquire unique positional identities. In the second step, called cell type selection or commitment, daughter progenitor cells are produced from multipotent progenitors. This newly produced progenitor cells can only give rise to a single neural cell type which may be neurons, oligodendrocytes or astrocytes. In the next step of subtype specification, the committed neuronal progenitors become restricted to produce particular kinds of neurons such as motor neurons or interneurons. Once these particular neurons are produced, the neuronal progenitors become postmitotic and start to migrate to more differentiated areas of the neural tube to terminally differentiate. Oligodendrocytes, astrocytes and some types of neurons begin to migrate and differentiate while they are still dividing.

During neuronal differentiation, ventricular zone progenitors transit from neurogenic to gliogenic divisions, and this process is known as the gliogenic switch. While the molecular mechanisms that control this switch are not very well understood, the Notch signaling pathway seems to play a fundamental role by maintaining telencephalic progenitors in an undifferentiated state (Mizutani and Saito, 2005).

Besides Notch signalling pathway, JAK-STAT signaling in neural precursors was also shown to be effective in triggering gliogenesis (Barnabe-Heider et al., 2005; Ochiai et al., 2001).

NFI proteins role in neural development is underlined by their expression pattern. During embryogenesis, NFI genes are expressed in distinct, overlapping patterns with high levels of expression of Nfia, Nfib, and Nfix in the developing neocortex (Chaudhry et al., 1997). Initially, Nfia, Nfib, and Nfix are highly expressed in ventricular zones (Chaudhry et al., 1997; Deneen et al., 2006; Plachez et al., 2008). At late embryonic and early postnatal ages, expression is particularly high in the cerebral cortex, hippocampus, basilar pontine nuclei, cerebellum, and spinal cord (Chaudhry et al., 1997; Gesemann et al., 2001; Shu et al., 2003; Deneen et al., 2006; Wang et al., 2007; Plachez et al., 2008).

The role of NFI proteins in gliogenesis has been well established in several regions of the CNS (Namihira et al., 2009; Deneen et al., 2006; Piper et al., 2010; Kang et al., 2012). Previous loss of function studies (das Neves et al., 1999; Shu et al., 2003; Steele-Perkins et al., 2005; Deneen et al., 2006; Driller et al., 2007; Wang et al., 2007; Campbell et al., 2008) show that in order for the neuronal and glial populations to properly develop in cerebral cortex, hippocampus, cerebellum, and spinal cord, NFI proteins are required.

Nfia acts downstream from Notch to facilitate expression of astrocyte specific genes, such as GFAP and astrocyte differentiation in cortical cultures as the astrocytic gene promoters were protected from DNMT1 (DNA (Cytosine-5-)-Methyltransferase 1) via the binding of NFIA to these promoters (Namihira et al., 2009). However, in the embryonic spinal cord, Notch signaling during the gliogenic switch did not result in the induction of NFIA or gliogenesis (Deneen et al., 2006). *Nfia* also downregulates the activity of the Notch signaling pathway via repression of the key Notch effector *Hes1* (Piper et al., 2010).

Another role of Nfia in gliogenesis is that it interacts with Sox9 to arrest neurogenesis and regulate genes associated with gliogenesis (Kang et al., 2012). *Nfia* and *Nfib* also modulates gliogenesis and oligodendrocyte differentiation within the developing spinal cord (Deneen et al., 2006) and hippocampus (Piper et al., 2010).

The lack of midline glia and the abnormalities in protein expression lead to agenesis of corpus collosum (Shu et al., 2003; das Neves et al., 1999). Moreover, *Nfib* regulates expression of guidance molecules such as Slit1 and neuropilin leading to defects in midline crossing of commissural axons and agenesis of corpus collosum (Piper et al., 2009). *Nfib*$^{-/-}$ mice exhibits delayed glial differentiation and maturation within the developing neocortex and hippocampus (Barry et al., 2008; Piper et al., 2009), supporting Nfib's role in regulation of glial development in the CNS.

These in vivo data are complemented by in vitro studies on human embryonic stem cells which show that while NFI-A and NFI-B are expressed during early gliogenesis, NFI-X and NFI-C are expressed later in the differentiation of astrocytes and control the expression of late astrocyte markers including GFAP and SPARCL1 (Wilczynska et al., 2009).

Interestingly, while they regulate gliogenesis in the developing neocortex, hippocampus, and spinal cord, NFI proteins also regulate cell migration and terminal differentiation in the cerebellum and glia mediated cell migration in the hippocampus (Barry et al., 2008; Wang et al., 2007).

1.1.5 Roles of NFI family in cancer and neurodegenerative diseases

NFI family members are not only involved in developmental processes but also in neurological diseases and cancer.

According to Lu et al. (Lu et al., 2007), NFIA haploinsufficiency resulted in central nervous system malformation such as tethered spinal cord, Chiari type I malformation, seizures, hypoplastic or absent corpus callosum, and hydrocephalus. Additionally, Koehler et al. (2010) reported that deletion of a 4.9 Mb part of NFIA gene resulted in hypoplasia of the corpus callosum, ventriculomegaly, and dysmorphic features. Also, NFIB has been shown to mediate the function of a transcriptional activator that is associated with Autism Spectrum Disorder (Choi et al., 2011).

In addition to NFIA and NFIB's involvement in human neurological conditions, they have also been implicated in cancer. Using genomic hybridization and DNA sequencing NFI-A gene was shown to be associated with different chronic malignant myeloid diseases with structural alterations of the NFI-A gene (Bernard et al., 2009). Similarly, NFIB's has been shown to translocate with high-mobility group

(nonhistone chromosomal) protein isoform I-C (HMGIC) in pleomorphic adenomas has been shown (Geurts et al., 1998).

Conversely, when NFI proteins were overexpressed in chick embryo fibroblasts cells, they cause fibroblasts cells to become more resistant to transformation by a several nuclear oncogenes, including fos, jun and qin (Schuur et al., 1995). In a recent study (Moon et al., 2011), the expression frequency of NFIB protein in human breast cancer tissues and its clinical implications were investigated and significant level of NFIB transcripts were seen in triple negative breast cancer. Additionally, silencing the NFIB gene lead to significant reduction in cell proliferation and increase in apoptotic signaling pathway.

Another study implicating NFIB in cancer showed that the fusion of MYB oncogene and NFIB is common in Adenoid cystic carcinoma (ACC) (Persson et al., 2009) leading to the overexpression of MYB in the salivary gland (Mitani et al., 2010). Interestingly, Fujita et al. (2008) predicted and confirmed that the NFIB mRNA is a target of miR-21, a key microRNA which is known to be highly expressed in various cancers (Selcuklu et al., 2009).

1.2 Experimental Background

1.2.1 Screening technologies for protein-protein interactions

All cellular processes involve protein-protein interactions. For analyzing protein-protein interactions, several methods have been developed so far (reviewed in Brückner et al., 2009). Some of these are copurification, affinity purification or coimmunoprecipitation of protein complexes. These techniques test for interactions in vitro. Mass Spectrometry (MS) coupled to affinity purification techniques can be used to identify interacting partners of a specific protein. Although this approach is straightforward, it is costly and time-consuming.

Phage display or yeast two-hybrid (Y2H) techniques can also be good candidates to screen for potential binding partners of particular proteins. These techniques make use of cDNA-expression libraries. The cDNA library is expressed in bacteriophage with the bacteriophage coat protein on the surface of the viral particle which is then allowed to bind to bait protein immobilized on a solid surface. Attached phages can then be isolated and characterized.

Y2H and affinity purification coupled to MS (AP/MS) are preferred methods for automated screening of interactions (Brückner et al., 2009). In the case of MS, mass-to-charge ratio of ionized molecules are determined to identify proteolytic fragments of proteins or even entire proteins and protein complexes. However, MS requires expensive and large equipment. This is a disadvantage when compared to Y2H. Consistently, it is estimated that half of the interaction data available on databases such as IntAct and MINT arise from Y2H assays (Brückner et al., 2009).

1.2.1.1 Yeast two hybrid system

Originally created by Fields and Song (Fields and Song, 1989), yeast two-hybrid technique has been used extensively to search for interacting proteins. In the Y2H system, the interaction between two proteins of interest is detected via the reconstitution of a transcription factor and the subsequent activation of reporter genes under the control of this transcription factor (Auerbach et al., 2005).

To analyze interactions between two proteins, GAL-4 DNA binding domain (GAL4-DBD) and GAL-4 activation domain (GAL4-AD) are used to make fusion protein constructs. For constructing bait, the protein of interest is fused to N-terminal part of GAL4-DBD and for constructing prey, the interactor protein is fused to the C-terminal part of GAL4-AD. When these constructs are transformed into yeast and fusion proteins are expressed, the two separate polypeptides reconstitute Gal4 transcription factor, subsequent recruitment of RNA polymerase II by Gal4 leads to transcription of a GAL1-lacZ fusion gene. Since this reporter gene encodes the beta-galactosidase enzyme, the interaction can be checked with a colorimetric assay.

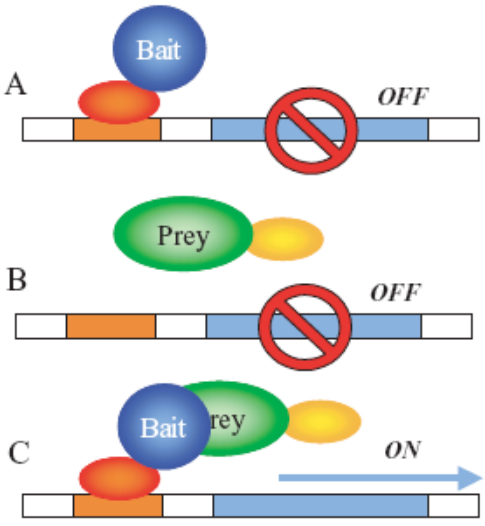


Figure 1.3 The yeast two-hybrid system. (A) A bait is expressed as a fusion to a DNA binding domain (DBD). The DBD–bait binds to the operator sequences present in the promoter region upstream of the reporter gene but does not activate its transcription since the DBD–bait does not contain an activation domain. (B) A prey is expressed as a fusion to an activation domain (AD). The AD–prey fusion has the capability to activate transcription in yeast but because it is not actively targeted to the promoter it does not activate transcription of the reporter gene. (C) The interaction between bait and prey targets the AD–prey

fusion protein to the promoter, thereby reconstituting an active transcription factor. The hybrid transcription factor is bound to the promoter upstream of the reporter gene and therefore activates transcription (Auerbach et al., 2005).

In the Matchmaker GAL4-based two-hybrid assay, a bait protein is fused to the Gal4 DNA-binding domain (DNA-BD), while libraries of prey proteins are fused to the Gal4 activation domain (AD). In the Matchmaker Gold Yeast Two-Hybrid System, when bait and library (prey) fusion proteins interact, the DNA-BD and AD are brought into proximity to activate transcription of four independent reporter genes (AUR1-C, ADE2, HIS3, and MEL1).

There are four integrated reporter genes under the control of three distinct Gal4-responsive promoters (Figure 1.4) in Clontech's Y2HGold Yeast Strain, which are used to detect two-hybrid interactions:

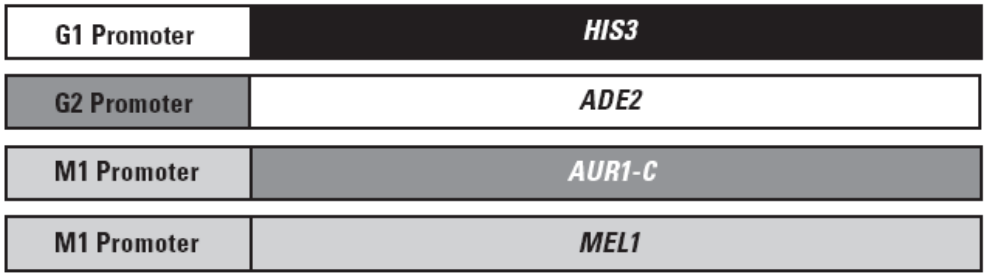
AUR1-C. The reporter gene AUR1-C encodes the enzyme inositol phosphoryl ceramide synthase. This gene is expressed in *Saccharomyces cerevisiae* as a result of protein-protein interaction. The enzyme expressed by AUR1-C then confers resistance to the toxic drug Aureobasidin A.

HIS3. Y2HGold strain is unable to grow on media that lack of histidine, as it is unable to synthesize this amino acid. When an interaction occurs between bait and prey proteins, this reporter gene is expressed and allows cell to biosynthesize histidine and to grow on minimal medium which lacks of histidine.

ADE2. Y2HGold strain is unable to grow on media that lack of adenine, as it is unable to synthesize this amino acid. When an interaction occurs between bait and prey proteins, this reporter gene is expressed and allows cell to biosynthesize adenine and to grow on minimal medium which lacks of adenine.

MEL1. This reporter gene encodes the enzyme α-galactosidase. When a two-hybrid interaction occurs between two proteins, α-galactosidase is expressed and digests chromagenic substrate X-α-Gal resulting in blue color on medium.

Y2HGold (Mating Partner) reporter gene constructs



Y187 (Library Host Strain) reporter gene constructs

Figure 1.4 Reporters of Y2HGold and Y187 strains (Clontech, PT4084-1 (PR033493) Published 30 March 2010).

1.3 NFI Binding Proteins

NFI proteins have been shown to interact with several proteins and in many cases, this interaction was identified by yeast two hybrid technique. Some of these protein interaction based studies are described below.

Yeast two hybrid technique was used for identifying proteins that may interact with proline-rich transactivation sequence of Nfic which contains a TGF-β responsive domain (TRD) (Alevizopoulos et al., 1995). The interacting protein was shown to be histone H3, point mutations which were made in the TRD reduced TGF-β induction and Nfic binding to histone H3. These interactions may act downstream and regulate chromatin structure in response to TGF-β signaling.

Another study (Ferrari et al., 2004) showed that Nfic interacts with histone H3 to mediate the boundary effect that protects genes near the telomeres from silencing. A deletion mutagenesis localized the barrier activity to the Nfic histone-binding domain and identified several amino acids that were involved in the boundary and histone binding activities.

A mammalian two hybrid study (Bachurski et al., 2003) concerning Surfactant protein C(SP-C) and Thyroid transcription factor 1 (TTF-1) was performed. The proximal promotor region of SP-C, contains a binding site for NFI and TTF-1. To understand whether or not NFI isoforms regulate SP-C transcription by interacting

with TTF-1, NFI family members and TTF-1 cotransfected into JEG-3 cells. A synergistic activation was seen on SP-C promoter. By doing coimmunoprecipitation and mammalian two hybrid assays, it was shown that TTF-1 directly interacts with conserved DNA binding and dimerization domains of all NFI family members. Interaction between these proteins activates SP-C transcription.

NFI proteins may negatively regulate transcription of phosphoenolpyruvate carboxykinase (GTP) (PEPCK) by competing with cAMP regulatory element-binding protein (CREB) for the CREB binding site on CBP (CREB binding protein) (Leahy et al., 1999). To demonstrate direct interaction between the transactivation domain of NFI-C (220–499) and the CREB binding domain of the CREB-binding protein (CBP) a mammalian two-hybrid system was used.

Tanese et al. (1991) showed that a Gal4 fusion protein that carries the proline rich activation domain of NFIC (carboxy-terminal 100 amino acids) can interact with purified coactivator proteins to initiate transcription in vitro.

A yeast model system was utilized to study the functional interaction between the human estrogen receptor (hER) and NFIC (Tsai-Pflugfelder et al., 1998). Using chimeric proteins of NFIC and GAL4 to measure activation of an ER- and CTF1-dependent reporter plasmids, they showed that proline-rich transactivation domain of NFIC is responsible for this interaction.

Xiao et al. (1994) found that the proline-rich transcriptional activation domain of the NFIC contains a sequence with striking similarity to the heptapeptide repeats of the CTD and this CTD-like motif is essential for the transcriptional activator function of the proline-rich domain of NFIC. Additionally, the proline-rich activation domain of NFIC interacts directly with the TATA-box-binding protein (TBP), and a mutation in the CTD-like motif that abolished transcriptional activation reduces the affinity of the proline-rich domain for TBP which was shown by protein affinity chromatography.

Through a Y2H screen, NFIX was shown (Singh et al., 2009) to interact with high mobility group protein (HMGN1) and the CGG triplet repeat binding protein 1 (CGGBP1) and these interactions are impotant in regulation of HSF1 (a heat shock-related transcription factor) expression. In this Y2H experiment, CTF1 domain of NFIX was selected as bait and a fetal human brain cDNA library was screened. Upon

a heat shock, the transcription of HSF1 is inhibited by a protein complex contaning NFIX, CGGBP1 and HMGN1. This requires CGG repeat in HSF1 promoter and HSF1 binding site in NFIX promoter.

1.4 Aim of the Study

NFI family of transcriptional factors has four members: Nfia, Nfib, Nfic, and Nfix. These genes are expressed in specific overlapping patterns during fetal development resulting in more than 9 isoforms due to alternative splicing. The typical NFI protein is composed of an N-terminal DNA-binding/dimerization domain and C-terminal transcriptional activation and/or repression domains (Altmann et al., 1994; Chaudhry et al., 1997; Kruse et al., 1991; Kruse and Sippel, 1994a; Santoro et al., 1988; Paonessa et al., 1988).

NFI proteins act as cellular transcription/replication factors and play a key role in central nervous system development, including cell migration, terminal differentiation as well as neurogenesis and gliogenesis. These proteins are required for proper development of neuronal and glial populations in the cerebral cortex, hippocampus, cerebellum, and spinal cord (Barry et al., 2008; Wang et al., 2007).

The members of NFI family regulate different developmental processes and genes in different parts of the brain. For example, in cerebellum, their roles are directed towards migration and terminal differentiation, wheras in forebrain, they are mostly involved in neurogenesis and gliogenesis, and in spinal cord, they regulate gliogenesis and astrocyte differentiation (Deneen et al., 2006; Piper et al., 2010).

Since the DNA-Binding domains of NFI proteins are highly conserved and their affinities for consensus DNA binding sites are very similar, the variation between their roles in transcriptional regulation can be explained by “combinatorial control”. Combination of several transcription factors and chromatin modifying proteins lead to the final readout of gene expression. Thus, if NFI proteins interact with different proteins in different CNS regions and cell types during development, they will have different effects on the transcriptional regulation and as a result, they will play different roles in forebrain, neocortex, hippocampus and cerebellum.

The fact that knockout of NFI genes that are expressed in the same brain region can have different phenotypes and NFI proteins can regulate different processes and

genes in different brain regions suggests that some of the interaction partners of NFI proteins in CNS still need to be identified to better understand their roles in transcriptional regulation and in brain diseases.

In this study, we performed a Yeast Two-Hybrid Screen to search for interactors of NFIB from human fetal brain cDNA library. Although the role of NFI family members, in particular NFIB protein, in the developing CNS has been studied extensively, there is no report of a screen for interacting partners of NFIB during neural development.

Since NFIB is a transcription factor, it could autoactivate related reporters used during two hybrid screening. To avoid such false positive interactions, in addition to full NFIB protein, we also made a prey construct of truncated NFIB where its DNA-binding domain is missing. The findings of this study will reveal unidentified interactions which may be important in brain development and diseases.

2. MATERIALS AND METHODS

2.1 Materials

2.1.1. Equipment

The equipment that was used in this thesis project is listed below:

Table 2.1 Equipment.

Equipment	Brand & Supplier Company
DNA Sequencer	Applied Biosystems
Gel Electrophoresis System	Bio-Rad
Microcentrifuge	ScanSpeed
pH Meter	WTW
Precision Weigher	Precisa
Shaker	Sartorius
Rocking platform shaker	Heidolph
Vortex	LMS Group
Water Bath	Memmert
Thermal Cycler	Applied Biosystems
Nucleic acid Spectrometer	NanoDrop/Thermoscientific
Tabletop Centrifuge	Beckman Coulter

2.1.2 Chemicals

The chemicals that were used in this thesis project are shown below:

Table 2.2 Chemicals.

Chemical	Supplier Company
Agar, Bacteriological	Acumedia
Low Electroendosmosis Agarose	Lonza
Yeast Nitrogen Base	Clontech
-Leu/-Trp Drop Out Amino Acid Mixture	Clontech
-Leu Drop Out Amino Acid Mixture	Clontech
-Trp Drop Out Amino Acid Mixture	Clontech
-Ade/-His/-Leu/-Trp/ Drop Out Amino Acid Mixture	Clontech

Table 2.2 (continued) : Chemicals.

Chemical	Supplier Company
Isopropanol	J.T.Baker
Ethanol, absolute	Carlo Erba
Methanol,analytical grade	Fisher Scientific
Acetic Acid	J.T.Baker
$MgCl_2$	Carlo Erba
NaCl	Merck
KH_2PO_4	Merck
Dextrose	Amresco
Peptone G	Acumedia
Glycine	Fischer Scientific
Tryptone	Lab M
Yeast Extract	Lab M
Tris Base	Fischer Scientific
EDTA	Carlo Erba
Adenine hemisulfate salt	Sigma
PronaSafe Nucleic Acid Staining Dye	Conda Lab
Ethidium Bromide	Fischer Scientific
X-α-Gal	Clontech

2.1.3 Reagents

The reagents including primers, enzymes and buffers that were used in this thesis project are listed below;

Table 2.3 Reagents.

Reagent	Brand & Supplier Company
1 kb DNA Ladder	Fermentas
6X DNA Loading Dye	Fermentas
10 mM dNTP Mix	Fermentas
OneTaq® DNA Polymerase	New England Biolabs
5X OneTaq Standard Reaction Buffer	New England Biolabs
XmaI Restriction Enzyme	New England Biolabs
Sal I Restriction Enzyme	New England Biolabs
Hae III Restriction Enzyme	Thermo Scientific
10X Fast Digest Buffer	Thermo Scientific
NEBuffer	New England Biolabs
Lyticase from *Anthrobacter luteus*	Sigma Aldrich
50X Advantage 2 polymerase mix	Clontech
10X Advantage 2 PCR Buffer	Clontech

Table 2.3 (continued) : Reagents.

Reagent	Brand & Supplier Company
α-NFI primary antibody	Santa Cruz Biotechnology
α-myc primary antibody	Clontech
GAL4 DNA-BD monoclonal antibody	Clontech
Goat anti-mouse seconder antibody	Thermo Scientific
Goat anti-rabbit seconder antibody	Thermo Scientific

2.1.4 Commercial kits

The commercial kits that were used in this thesis project are shown below:

Table 2.4 Commercial kits.

Kit	Brand & Supplier Company
High Pure Plasmid Purification Kit	Roche Applied Science
Agarose Gel DNA Extraction Kit	Roche Applied Science
Yeastmaker Yeast Transformation System 2	Clontech
Pretransformed Human Fetal Brain cDNA Library	Clontech
MatchMaker Gold Yeast Two-Hybrid System	Clontech
In Fusion HD Cloning Kit	Clontech

2.1.5 Buffers and solutions

The buffers and solutions that were used in this thesis project are shown below:

Table 2.5 Buffers and solutions.

Buffer/Solution	Content
TAE Buffer (50X)	40mM Tris, 20 mM Acetic Acid, 1 mM EDTA(pH:8.0), dH_2O
1.1X TE/LiAc	1.1X LiAc, 1.1X TE, dH_2O
PEG/LiAc	40% (w/v) PEG 3350, 1X TE buffer, 1X LiAC
0.9% NaCl	0.9 g NaCl, 100 ml dH_2O

2.1.6 Microorganism strains

2.1.6.1 Bacterial strains

Escherichia coli (*E.coli*) XL1-Blue strain endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB+ lacIq Δ(lacZ)M15] hsdR17(rK- mK+)

2.1.6.2 Yeast strains

- *Saccharomyces cerevisiae* Y2HGold strain (Clontech):

MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2 : : $GAL1_{UAS}$–$Gal1_{TATA}$–His3, $GAL2_{UAS}$–$Gal2_{TATA}$–Ade2URA3 : : $MEL1_{UAS}$–$Mel1_{TATA}$ AUR1-C MEL1

- Saccharomyces cerevisiae Y187 strain (Clontech):

MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, gal80Δ, met–, URA3 : : $GAL1_{UAS}$–$Gal1_{TATA}$–LacZ, MEL1

2.1.7 Culture media

LB Medium: 10 g tryptone, 5 g yeast extract, and 10 g NaCl were dissolved in ultrapure water and the volume was adjusted to 1 L. For sterilization, the media was autoclaved at 121 ^{o}C for 15 minutes. For preparing LB agar plates, 15 g agar was added to 1 L LB medium before sterilization. For preparing selective medium, the medium was allowed to cool to 55 ^{o}C following sterilization, and then the relevant antibiotics were added. The stock and working concentrations of antibiotics used in this thesis are listed in Table 2.6.

Table 2.6 Antibiotics and their concentrations.

Antibiotic	Stock Solution Concentraion	Working Solution Concentraion
Kanamycin	10 mg/L in ultrapure water	50 µg/L in medium
Ampicillin	10 mg/L in ultrapure water	100 µg/L in medium

SOB Medium: 2 g tryptone, 0.5 g yeast extract, 0.058 g NaCl, 0.0186 g KCl were dissolved in 100 ml ultrapure water and autoclaved at 121 ^{o}C for 15 minutes.

SOC Medium: SOC medium was prepared by adding 0.095 g $MgCl_2$, and 0.24 g $MgSO_4$, and 0.36 g glucose into SOB medium followed by filter sterilization (0.22 µm).

HTB Solution: 10 mM HEPES, 15 mM $CaCl_2$, 250 mM KCl were dissolved in dH_2O. pH of the solution was adjusted to 6.7 with 1 M KOH. $MnCl_2$ was added to a final concentration of 55 mM and volume was brought to 200 ml. The solution was filter sterilized (0.45µM).

YPD medium: 20 g Peptone G, 10 g Yeast Extract, and 2 g Dextrose were dissolved in 1 L ultrapure water followed by autoclaving at 121 ^{o}C for 15 minutes. For preparing YPD agar plates, agar was added to YPD medium at 20 g/L before sterilization.

YPDA medium: YPDA medium is adenine supplemented version of YPD medium, Adenine hemisulfate was added to YPD medium after sterilization at a working concentration of 120 mg/L.

SD Medium: 20 g dextrose and 6.7 g yeast nitrogen base without amino acids, were dissolved in 1 L ultrapure water to prepare SD base medium. The appropriate amount of drop-out amino acid mixture shown in Table 2.7 was then added to this SD base and the resulting solution was autoclaved at 121 ^{o}C for 15 minutes. For SD agar plates, agar was added to a final concentration of 20 g/L. The drop-out amino acid mixtures with their final concentrations are listed below:

Table 2.7 Concentrations of drop-out (DO) nutrients.

Nutrient	Final Concentration
-Trp DO Supplement	0.74 g/L
-Leu DO Supplement	0.69 g/L
-Leu/-Trp DO Supplement	0.64 g/L
-Ade/-His/-Leu/-Trp DO Supplement	0.60 g/L

2.1.8 Vectors

2.1.8.1 pGBKT7, DNA binding domain vector

The proteins that are expressed from pGBKT7 vector are fused to amino acids 1–147 of the GAL4 DNA binding domain (DNA-BD). The pGBKT7 vector also contains the following main elements: a constitutive ADH1 promoter (P_{ADH1}) from which fusion proteins are expressed at high levels; T7 and ADH1 termination signals ($T_{T7\ \&\ ADH1}$) which terminates transcription; a T7 promoter; a c-Myc epitope tag; a MCS; pUC and 2 μ ori from which pGBKT7 replicates autonomously in both *E. coli* and *S. cerevisiae*, respectively; kanamycin resistance gene (Kan^{r}) for selection in *E. Coli*; TRP1 nutritional marker for selection in yeast;

The restriction map and the multiple cloning site (MCS) of pGBKT7 along with other significant elements on this vector are shown in Figure 2.1 and 2.2.

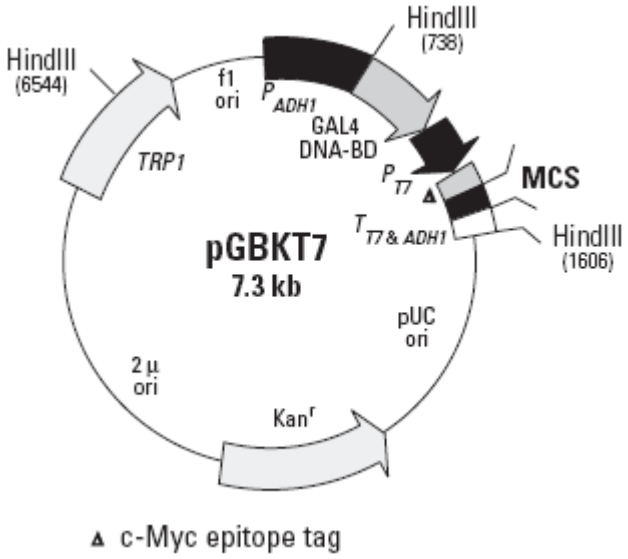


Figure 2.1 Map of pGBKT7 Vector

Figure 2.2 Restriction site of pGBKT7 vector

2.1.8.2 pGADT7, activation domain vector

The proteins that are expressed from pGADT7 vector are fused to amino acids 768–881 of the GAL4 activation domain (AD). The pGADT7 vector also contains the following main elements: a constitutive ADH1 promoter (P_{ADH1}) from which fusion proteins are expressed at high levels; T7 and ADH1 termination signals ($T_{T7\ \&\ ADH1}$) which terminates transcription; the SV40 nuclear localization sequences (added to the activation domain sequence) by which the fusion protein is targeted to the yeast nucleus; a T7 promoter; an HA epitope tag; a MCS; pUC and 2 µ ori from which

pGBKT7 replicates autonomously in both *E. coli* and *S. Cerevisiae*, respectively; an ampisillin resistance gene (Amp^r) for selection in *E. Coli*; LEU2 nutritional marker for selection in yeast.

The restriction map and the multiple cloning site (MCS) of pGADT7 along with other important elements on vector are shown in Figure 2.3 and 2.4.

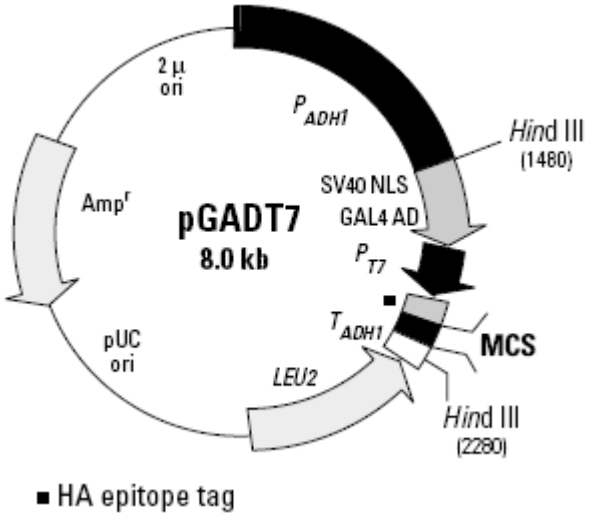


Figure 2.3 Map of pGADT7 vector

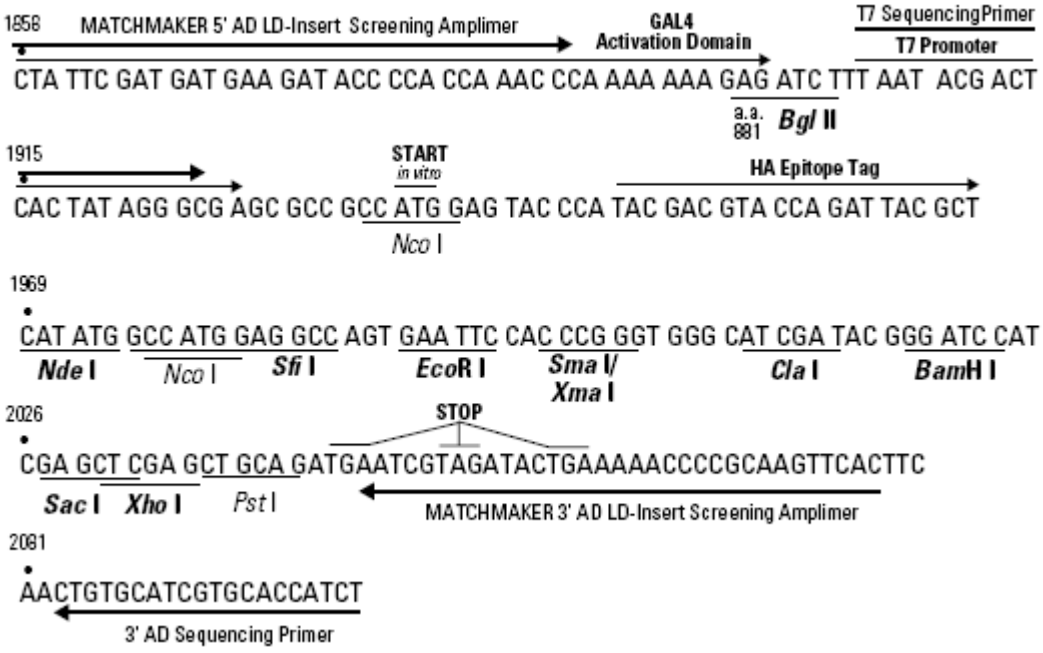


Figure 2.4 Restriction site of pGADT7 vector

2.1.8.3 pCHNFI-B2, insert vector

Nfib3 cDNA is cloned in pCHNFI-B2 vector. This construct was kindly provided by Richard Gronostajski (University at Buffalo, New York, USA).

This vector contains 2238 bp long mouse Nfib3 cDNA which codes for a 435 aa protein fused to a N-terminal hemagglutinin epitope tag. The construction of pCHNFI-B2 plasmid has been described previously (Chaudhry et al.,1997).

2.1.9 Primers

The primers used in this thesis were as follows;

Table 2.8 Primers

Primer	Sequence	Company
mNFIB2_Xma_IF5	AGGCCGAATTCCCGGGTATGATGTAT TCTCCCATCTGTCTC	Alpha DNA
mNFIB2_Xma_IF6	AGGCCGAATTCCCGGGTGAGGACAG CTTTGTAAAATCCGG	Alpha DNA
mNFIB2_Sal_IFR	GCCGCTGCAGGTCGACTCTAGCCCAG GTACCAGGA	Alpha DNA
3' DNA-BD Sequencing Primer	TTTTCGTTTTAAAACCTAAGAGTC	Alpha DNA
T7 Sequencing Primer	TAATACGACTCACTATAGGGCG	Iontek
5'AD-LD-ISA	CTATTCGATGATGAAGATACCCCACC AAACCC	Alpha DNA
3' AD-LD-ISA	GTGAACTTGCGGGGTTTTTCAGTATC TACGAT	Alpha DNA

2.2 Methods

2.2.1 Construction of bait plasmid

Mouse NFIB2 cDNA and truncated version of NFIB2 cDNA (NFIB2Δ) was cloned into pGBKT7 expression vector by PCR amplifying it from pCHNFI-B2 construct using the following primer pairings: mNFIB2_Xma_IF5,-mNFIB2_Sal_IFR and mNFIB2_Xma_IF6- mNFIB2_Sal_IFR.

2.2.1.1 Restriction digestion of pGBKT7 expression vector

For cloning NFIB2 inserts into pGBKT7 vector, pGBKT7 was restriction digested with SalI and XmaI 6-cutter enzymes whose digestion patterns are shown below;

XmaI

```
5′-  C/CCGGG  -3′
3′-  GGGCC/C  -5′
```

SalI

```
5′-  G/TCGAC -3′
3′-  CAGCT/G -5′
```

Figure 2.5 Digestion patterns of restriction enzymes

Partial sequence of the multiple cloning site of pGBKT7 after restriction digestion is shown below. The shaded bases show recognition sites of restriction enzymes.

```
5′…CATATGGCCATGGAGGCCGAATTC     -3′    CCGGGGATCCG       5′- T CGA CCTGCAGCGGCCGCATAA…3′
3′…GTATACCGGTACCTCCGGCTTAAG GGC C -5′     CCTAGGCAGCT    3′-       GGACGTCGCCGGCGTATT…5′
```

Figure 2.6 Multiple cloning site of pGBKT7 vector

The restriction was carried out at 37ºC, overnight. At the end of the restriction reaction, a heat shock was applied at 65 ºC for 20 minutes in a water bath. The reaction components were indicated in Table 2.9.

Table 2.9 Components of restriction reaction.

Reaction Components	Amounts / Volumes
dH_2O	40.3 µl
NEB4	5 µl
pGBKT7 Vector	3.25 µl : 5 µg
SalI	0.5 µl
XmaI	0.5 µl
TOTAL	**50 µl**

2.2.1.2 Primers for amplification of pCHNFI-B2 vector

To amplify NFIB2 cDNA and NFIB2Δ cDNA from pCHNFI-B2 vector, primers were designed according to the instructions in the Clontech's InFusion Cloning System.

First, the primers contained 15 bases of homology (Figure 2.6; underlined bases) with the ends of the linearized vector. The 5' overhangs generated on the linearized pGBKT7 vector through restriction digestion were included in this 15 bp sequence. Secondly, one additional nucleotide was added to each primer in order to reconstruct XmaI and SalI restriction sites. These additional nucleotides were not part of the 15 bases of sequence homology. Thirdly, appropriate number of bases were selected on extreme sequences of NFIB2 and NFIB2Δ for implementing general primer design rules. Lastly, an extra base was added adjacent to the reconstructed restriction sites in the primer for 'in frame' cloning.

Below is the primer backbones that represent the ones used in actual experiment.

5'- $\mathbf{V_{15}}$ + $\mathbf{R_S}$ + $\mathbf{I_F}$ + $\mathbf{N_n}$ -3'

Figure 2.7 A representation of the primers used (**V_{15}**: Sequence specific to pGBKT7 ***V***ector (15 bases), **R_S**: Additional base for reconstructing **R**estriction **S**ites, **I_F:** Additional base for '**I**n **F**rame' cloning, **N_N**: ***n*** bases specific to ***N***FIB2 cDNA)

In the figures below, representative sequences of NFIB2 and NFIB2Δ are shown;

```
5'- ATGATGTATTCTCCCATCTGTCTCACTCAGGATG……CTCAAACGAGCCAGTCCTGGTACCTGGGCTAG -3'
3'- TACTACATAAGAGGGTAGACAGAGTGAGTCCTAC……GAGTTTGCTCGGTCAGGACCATGGACCCGATC -5'
```

Figure 2.8a The extreme sequences of NFIB2

```
5'- GAGGACAGCTTTGTAAAATCCGGAGTCTTCAATGTA……CAAACGAGCCAGTCCTGGTACCTGGGCTAG -3'
3'- CTCCTGTCGAAACATTTTAGGCCTCAGAAGTTACAT……GTTTGCTCGGTCAGGACCATGGACCCGATC -5'
```

Figure 2.8b The extreme sequences of NFIB2Δ

In the figures above, the underlined bases are the ones that were included in primers. The sequence specific to the NFIB2 cDNA or NFIB2Δ cDNA must be 18–25 bases long, with %GC: 40-60, T_m: 58-65°C, $Tm_{|G1-G2|} \leq 4°C$.

For designing primers, we also utilized Clontech's online tool which can be accessed via the following link: http://bioinfo.clontech.com/infusion/

2.2.1.3 Amplification of NFIB2 inserts by PCR

A PCR reaction was setup to amplify the NFIB2 and NFIB2Δ fragments using the primers described above.

Below are the reaction components to carry out this PCR reaction;

Table 2.10 Components in the PCR reaction for pCHNFIB2 vector.

Components	Volumes/Amounts
dH_2O	16.375 µl
5X One Taq Standart Reaction Buffer	5 µl
10 mM dNTPs	0.5 µl
10 µm Forward Primer	0.5 µl
10 µm Reverse Primer	0.5 µl
One Taq DNA Polymrease	0.125 µl
TOTAL	**23 µl**

12.5 ng pCHNFI-B2 was added to the 23 µl master mix. The reactions were carried out in quintuplicate. At the end of amplification, the PCR products were combined for purification. The cycling protocol for this PCR reaction was given in Figure 2.9.

94 ºC : 1 minutes

- 94 ºC : 30 seconds
- 54 ºC : 30 seconds
- 68 ºC : 90 seconds

25 cycles

68 ºC : 5 minutes

Figure 2.9 PCR cycling protocol for amplification of NFIB2 inserts

2.2.1.4 Purification of NFIB2 insert

The PCR products were resolved on agarose gel and resulting bands were purified by Agarose Gel DNA Extraction Kit, Roche according to the manufacturer's instruction. Briefly;

- 300 µl Binding Buffer per 100 mg gel was added in a sterile microcentrifuge tube and the tube was vortexed for 15-30 seconds.
- The suspension was incubated for 10 minutes at 56 ºC by vortexing in every 2-3 minutes.
- 150 µl isopropanol per 100 mg gel was added and vortexed.
- Filter tube was placed in a collection tube and the suspension was transferred into filter tube.
- The suspension was centrifuged for 30-60 seconds at maximum speed.
- The flow-through was discarded and the filter tube was placed in the same collection tube.
- 500 µl wash buffer was added and the tube was centrifuged for 1 minutes at maximum speed.
- The flow-through was discarded and the filter tube was placed in the same collection tube.
- 200 µl wash buffer was added and the tube was centrifuged for 1 minutes at maximum speed.
- The flow-through and the collection tube were discarded. The filter tube was placed in a steril microcentrifuge tube.
- 50-100 µl elution buffer was added.
- The tube was centrifuged for 1 minutes at maximum speed.
- The tube was centrifuged for an additional 5 minutes at maximum speed.

- Without touching the base of the tube, the supernatant was transferred to a seperate tube. The purified DNA was in this supernatant.
- DNA concentration of purified DNA was measured by NanoDrop instrument.

2.2.1.5 Cloning NFIB2 cDNAs into pGBKT7 vector

For insertion of NFIB2 and NFIB2Δ amplicons into pGBKT7 vector separately, we used Clontech InFusion system. The reaction components of InFusion system and their amounts are shown below:

Table 2.11 Components in the InFusion reaction.

Reaction Components	**Amounts / Volumes**	
	NFIB2	**NFIB2Δ**
5X InFusion HD Enzyme Premix	2 µl	2 µl
Linearized Vector	100 ng: 1.85 µl	100 ng: 1.85 µl
Purified PCR Fragment	100 ng: 0.98 µl	100 ng: 2.14 µl
dH_2O	5.17 µl	4.01 µl
Total	**10 µl**	**10 µl**

The InFusion reaction was caried out at 50^{o}C for 15 minutes.

2.2.1.6 Transformation of InFusion cloning reaction

Following InFusion reaction, we transformed the resulting reaction mixtures into competent XL-1 Blue *Escherichia coli* cells. To prepare competent XL-1 Blue *E.coli* cells, we used competent cell protocol based on $CaCl_2$ treatment of *E.coli* cells. The competent cell protocol that we used was as follows:

- LB plate was streaked with cells from a frozen stock of XL-1 Blue.
- 10 fresh colonies were picked and transfered into 200 ml SOB liquid medium and incubated at room temperature until OD_{600} reached 0.45.
- The culture was divided into four 50 ml falcon tubes and kept on ice for 10 minutes.
- Falcon tubes were centrifuged at 2500 *g* for 15 minutes and supernatants were discarded.
- Cells were resuspended with 16ml HTB solution and kept on ice for 10 minutes.
- Falcon tubes were centrifuged at 2500 *g* for 15 minutes and supernatants were discarded.

- Cells were resuspended with 4ml HTB solution and tubes were combined.
- 1.2 ml DMSO was added and the tubes were gently mixed.
- The suspension was immediately aliquoted into microcentrifuges tubes each of which contained 200 µl cell suspension.
- Tubes were frozen rapidly using dry ice and placed in -80ºC.

Subsequent to preparing competent *E.coli* cells, we transformed them. The following protocol was applied for this:

- One fourth of InFusion reaction mixture (2.5 µl) was combined with 50 µl of competent cells in a sterile pre-chilled microcentrifuge tube.
- 1 µl of 40% DMSO was added to the tube and mixed.
- The tube was placed on ice bath for 30 minutes.
- A heat shock was applied at 42ºC for 45 seconds in a water bath.
- The tube was incubated on an ice bath for 2 minutes.
- SOC medium was added to the tube to a final volume of 500 µl.
- The transformation reaction was diluted 1/10 in 100 µl SOC medium and spread on a separate LB plate containing 50 µg/ml Kanamycin.
- The remainder of each transformation reaction was centrifuged at 6000 rpm for 5 min. Supernatant was discarded.
- Each pellet was resuspended in 100 µl fresh SOC medium. Each sample was spread on a separate LB plate containing 50 µg/ml Kanamycin.
- All of the plates were incubated overnight at 37°C.

2.2.1.7 Isolation of bait vectors

For isolating bait vectors, pGBKT7-NFIB2 and pGBKT7-NFIB2Δ constructs, we used Roche Plasmid Isolation Miniprep Kit.

The procedure of this kit is not much different from conventional plasmid purification kits. It starts with an alkaline lysis step which allows plasmid DNA to free from the cell. After centrifugation of cell suspension, choromosomal DNA is trapped in the cell debris and the supernatant is applied to the spin filter tube. The plasmid DNA is specifically adsorbed by the surface of glass fibers. A washing step ensures removal of the salts, proteins, and other cellular impurities from plasmid DNA. At the end, plasmid DNA is recovered with low salt elution buffer.

The procedure is given below;

- One colony was transferred into 3 ml of LB medium, with a yellow pipet tip in a 15 ml falcon tube and left for incubation at 37^{o}C overnight with 250 rpm shaking.
- Next day, culture was centrifuged at 3000 g for 5 minutes and supernatant was completely discarded.
- The pellet was suspended in 250 µl suspension buffer [50 mM Tris-HCl, 10 mM EDTA; pH: 8.0 at 25 ^{o}C] by pipetting.
- 250 µl lysis buffer [0.2 M NaOH, 1% SDS] was added to the tube and the suspension was mixed with inverting the tube 6 times.
- The suspension was incubated at room temperature for 5 minutes.
- 350 µl chilled binding buffer [4 M guanidine hydrochloride, 0.5 M potassium acetate; pH:4.2] was added and suspension was mixed with inverting the tube 6 times.
- The solution was incubated on ice for 5 minutes followed by centrifugation for 10 minutes at 13500 rpm.
- A filter tube was placed in a collection tube and supernatant was transferred into filter tube.
- The mixture was centrifuged for 1 minutes at maximum speed and flow-through was discarded.
- Filter tube was placed in the same collection tube and 500 µl Wash Buffer I [final concentrations after addition of ethanol: 5 M guanidine hydrochloride, 20 mM Tris-HCl, pH: 6.6 at 25 ^{o}C] was added to filter tube followed by 1 minute centrifugation at maximum speed.
- Filter tube was placed in the same collection tube and 700 µl Wash Buffer II [final concentrations after addition of ethanol: 20 mM NaCl, 2 mM Tris-HCl, pH: 7.5 at 25 ^{o}C] was added to filter tube followed by 1 minute centrifugation at maximum speed (13500 rpm).
- Flow-through was discarded and an additional minute centrifuge was performed to dry filter tube.
- Flow-through and collection tube were discarded and filter tube was placed in a fresh sterile microcentrifuge tube.

- 100 µl elution buffer [10 mM Tris-HCl; pH: 8.5 at 25 ^{o}C] was added to filter tube followed by 1 minute incubation step at room temperature.
- The tube was centrifuged for 1 minute at maximum speed and the concentration of plasmid DNA in the resulting eluate was calculated from A_{260} measurements with UV Spectroscopy.

2.2.1.8 Analytical restriction digestion of purified plasmids

To ascertain that purified plasmids contained the correct inserts, we digested the isolated plasmids with SalI which cuts pGBKT7, pGBKT7-NFIB2 and pGBKT7-NFIB2Δ at an only single recognition site. The reaction components were as follows;

Table 2.12 Components of restriction reaction.

Components	**Amount / Volume**	
	NFIB2	**NFIB2Δ**
dH_2O	19.63 µl	17.14 µl
10 X NEB4 Buffer	2.5 µl	2.5 µl
Plasmid DNA	2.37 µl : 500 ng	4.86 µl : 500 ng
Sal I restriction enzyme	0.5 µl : 10 U	0.5 µl : 10 U
TOTAL	**25 µl**	**25 µl**

The reactions were terminated the following day by heat shock at 65^{o}C for 20 minutes in a water bath. Subsequent to restriction digestion, one fifth of the reaction volumes were analyzed on a 0.8 % agarose gel.

2.2.1.9 Sequencing of pGBKT7-NFIB2 and pGBKT7-NFIB2Δ constructs

Selected purified plasmid DNAs were sent to sequencing to verify that the correct fragments were cloned into pGBKT7. The sequencing was performed by IonTek, Turkey.

The forward and reverse primers that were used for sequencing were chosen from upstream and downstream sites of NFIB2 and NFIB2Δ inserts. For this, 3' DNA-BD Sequencing Primer and T7 Sequencing Primer were used (indicated in Table 2.8).

2.2.2 Yeast transformation

After constructing the bait plasmids (pGBKT7-NFIB2 and pGBKT7-NFIB2Δ), we transformed them into the Y2HGold yeast strain to use in the subsequent mating

experiments. The following procedures, including competent cell preparation and transformation steps, were used in both bait strain transformation and in subsequent control experiments.

2.2.2.1 Preparation of competent cells

For competent cell preparation and subsequent transformation steps, Clontech Yeastmaker™ Yeast Transformation System 2 was used. This system is based on high-efficiency polyethylene glycol (PEG)/LiAc-based method for preparing and transforming competent yeast cells.

The following protocol was used for preparing competent yeast cells:

- A freshly prepared YPDA agar plate was streaked with yeast cells from a frozen yeast stock. The plate was incubated upside down at 30° C until colonies appear (3 days).
- One colony (diameter 2–3 mm, < 4 weeks old) was inoculated into 3 ml YPDA medium in a sterile 15 ml culture tube.
- The cell suspension was incubated at 30° C with shaking at 150 rpm for 10 hours.
- 5 µl of the culture was transferred to 50 ml of YPDA in a 250 ml flask.
- The culture was incubated with shaking at 150 rpm until the OD_{600} reaches 0.22.
- The cells were centrifuged at 700 *g* for 5 min at room temperature. The supernatant was discarded and the pellet was resuspended in 100 ml of fresh YPDA.
- The culture was incubated at 30° C until the OD_{600} reaches 0.45.
- 50 ml of the culture was transferred to a sterile falcon tube and the cells were centrifuged at 700 *g* for 5 min at room temperature. The supernatant was discarded and pellet was resuspended in 30 ml sterile, deionized H_20.
- The cells were centrifuged at 700 *g* for 5 min at room temperature. The supernatant was discared and pellet was resuspended in 1.5 ml of 1.1xTE/LiAc.
- The cell suspension was transferred to a 1.5 ml microcentrifuge tube and centrifuged at 13500 rpm for 15 sec.

- The supernatant was discarded and the pellet was resuspended in 600 µl of 1.1xTE/LiAc.

2.2.2.2 Transformation into Yeast

Once the competent cells were prepared, they were transformed with plasmid DNA.

The following protocol was applied for transformation:

- The Yeastmaker Carrier DNA (10 µg/µl) was denatured by heating to 95–100°C for 5 min, then cooled rapidly in an ice bath.
- 100 ng plasmid DNA and 10 µl carrier DNA were combined in a pre-chilled, sterile 1.5 ml tube.
- 50 µl of competent cells were added and the tube was gently mixed.
- 500 µl of PEG/LiAc was added and the tube was gently mixed.
- The suspension was incubated for 30 minutes at 30° C mixing cells every 10 minutes by tapping.
- 20 µl of DMSO was added and the tube was mixed.
- The tube was placed in a 42° C water bath for 15 minutes.
- Cells were mixed every 5 min by gently vortexing.
- The yeast cells were centrifuged at 13500 rpm for 15 seconds to obtain cell pellet.
- The supernatant was removed and the cells were resuspended in 1 ml YPD Plus Medium.
- Cells were incubated at 30° C for 1 hour with shaking.
- The yeast cells were centrifuged at 13500 rpm for 15 seconds to obtain cell pellet.
- The supernatant was discarded and cells were resuspended in 1 ml 0.9 % (w/v) NaCl Solution.
- 100 µl of 1/10 and 1/100 dilution was spread onto 100 mm plates containing the appropriate drop-out selection medium and X-α-Gal substrate.
- The plates were incubated upside down at 30° C until colonies appear (3 days).

2.2.3 Control experiments

Control experiments were performed before mating experiments. The control experiments consisted of four experiments: control for bait expression, control for bait toxicity, control for bait autoactivation, and control for mating.

2.2.3.1 Control for protein expression of bait vector

The bait vector must express NFIB protein in order for yeast two-hybrid screen to work. To look for yeast clones that express NFIB protein well, first we transformed pGBKT7-NFIB2 and pGBKT7-NFIB2Δ plasmids into Y2HGold competent cells, grew transformants on selection medium, extracted proteins from individual colonies, and subjected the extracts to Western Blot analysis.

Protein extraction from yeast

After preparing competent yeast cells and transforming them with the appropriate plasmids (pGBKT7-NFIB2 and pGBKT7-NFIB2Δ constructs), we grew transformed Y2HGold cells on selective medium and then extracted whole cell proteins by lysis. A number of colonies on selective plates were compared according their western bands in terms of their ability to express NFIB2 fusion protein.

The following protocol was used for growing the cultures and protein extraction:

- One colony (2-3 mm in diameter) was picked and transferred into 3ml SD/-Trp liquid drop-out medium in a 15 ml falcon tube and incubated at 30^{o}C overnight with shaking at 150 rpm.
- Next day, 5 ml fresh SD/-Trp medium was inoculated with the overnight culture to obtain a new culture with the initial OD_{600} value of 0.15.
- This new culture was incubated at 30^{o}C with shaking at 150 rpm until its OD_{600} reached 0.4-0.6.
- The cells were centrifuged at 4100 rpm for 15 minutes. The supernatant was discarded and cell pellet was suspended in 50 µl 1X sample buffer [0.0625M Tris-HCl: pH: 6.8, 10% (v/v) Glycerol, 2% SDS].
- This suspension was transferred to a clean microcentrifuge tube containing ≈50 µl glass beads, and mixed by pipetting.
- The resulting suspension was vortexed for 30 seconds and immediately placed on ice for 30 seconds. This was repeated 3 times.

- The resulting cell lysate was centrifuged at 16000 rpm for 5 minutes at 4^{o}C. The supernatant was transferred to a clean microcentrifuge tube.
- The tube was placed on a block heater and incubated at 95^{o}C for 5 minutes.
- The tube was centrifuged at 16000 rpm for 5 minutes at 4^{o}C and supernatant was transfered to a clean microcentrifuge tube.
- 25-40 µl of this protein extract was used in SDS-PAGE experiments.

SDS-PAGE electrophoresis and western blotting of protein extracts

Once the cell protein extracts were obtained, 30-40 µl of protein extracts were resolved on SDS-PAGE at 100 V for about 1.5 hours.

Following SDS-PAGE, the proteins were transferred from gel to the nitrocellulose membrane overnight at 15 V.

- The membrane was removed from transfer system and stained with Ponceau S solution.
- The presence of protein bands were confirmed visually, lanes and direction of blot were marked with a pencil.
- The membrane was soaked in 3% dry milk in TBST [50 mM Tris-HCl; pH: 7.5, 150 mM NaCl, 0.1% Tween 20] in a plastic bag and incubated for at least 30 minutes in room temperature with gentle agitation.
- Primary antibodies were diluted and the membrane was incubated with primary antibody solutions overnight at 4^{o}C with gentle agitation (2 hour at room temperature for Gal4-DNA-BD antibody). Dilution ratios and type of solutions shown in Table 2.13.
- The membrane was washed 4 times with TBST for 15 minutes at room temperature
- The secondary antibody was diluted and the membrane was incubated for 1 hour at room temperature with gentle agitation. Dilution rates and type of solutions are shown in Table 2.13.
- The membrane was washed 3 times with TBST and once with TBS [20 mM Tris-HCl; pH: 7.5, 140 mM NaCl] for 10 minutes at room temperature.

For western blotting and imaging of proteins, following protocol was used;

- The substrate solution [250 µl LumiGLO® Reagent A, 250 µl Peroxide Reagent B and 4.5 ml dH_2O] was prepared.
- The membrane was placed on a square piece of parafilm. Substrate reagent was applied onto it and incubated for 1 minute.
- After the membrane was removed from parafilm and cleaned from excess substrate solution, it was sandwiched in a piece of two layer stretch film, and then placed in a X-ray film cassette.
- One blue light sensitive autoradiography film was placed onto membrane and the cover of cassette was closed.
- The membrane was exposed to film for 1 minute. If required, a second piece of film was exposed for longer time periods (e.g. 5 minues). The film was removed from cassette and placed in autoradiography instrument.

Table 2.13 Dilution rates & solutions of antibodies.

Primary Antibody & dilution		**Secondary Antibody & dilution**	
α-myc	1/1000 in 3% dry	anti-mouse	1/500 in 3% dry milk
α-NFI	1/1000 in 3% dry	anti-rabbit	1/500 in 3% dry milk
Gal4-DNA-BD	0.5 ng/µl in TBST	anti-mouse	1/500 in TBST

2.2.3.2 Control for toxicity of bait vector

In order for the bait construct to be successfully used in the yeast two hybrid screen, the bait protein must not be toxic to yeast cells.

According to the instructions in the Clontech Matchmaker™ Yeast Two Hybrid System, if the colonies containing bait vector are significantly smaller than colonies containing the empty pGBKT7 vector, the expression of bait vector is considered to be toxic to the cells. Additionally, the cells containing toxic bait protein would grow on plates slower than the cells containing non-toxic bait protein.

The number and size of colonies resulting from transformation with either bait vectors or empty vector were compared.

2.2.3.3 Control for autoactivation of bait

The aim of autoactivation test is to confirm that the bait protein does not autonomously activate the reporter genes in Y2HGold, in the absence of a prey

protein. For this test, fresh 2-3mm colonies were streaked onto SD/Trp, SD/Trp/X-α-Gal, and SD/Trp/X-α-Gal/AbA plates. At the end of 3 days, the color of colonies were scored as blue, pale blue, or white.

2.2.3.4 Control for mating

Before doing actual mating and library screening experiments, a mating control was performed.

As a positive control, pGBKT7-53 that encodes the Gal4 DNA-BD fused to murine p53 was mated with pGADT7-T that encodes the Gal4 AD fused to SV40 large T-antigen. Since p53 and large T-antigen are known to interact in a yeast two-hybrid assay, mating Y2HGold [pGBKT7-53] with Y187 [pGADT7-T] will result in diploid cells containing both plasmids that can activate all four reporters.

A negative control was performed using pGBKT7-Lam (which encodes the Gal4 BD fused to lamin) and pGADT7-T. Diploid yeast containing pGBKT7-Lam and pGADT7-T should grow on SD/–Leu, SD/–Trp and SD/–Leu/–Trp (DDO) minimal media, but no colonies should grow on DDO + AbA plates.

For both positive and negative control matings the following protocol was used;

- Using the Yeast Transformation Protocol (small-scale) in the section 2.2.2, the following three transformations were done separately and plated on appropriate selective media: pGBKT7-53 and pGBKT7-Lam into Y2HGold plated on SD/-Trp, pGADT7-T into Y187 plated on SD/-Leu.
- Plates were incubated at 30^{o}C for 3 days and one 2–3 mm colony of each type was picked for use with the following small-scale mating procedure:
    - Positive Control Mating: Y2HGold [pGBKT7-53] and Y187 [pGADT7-T].
    - Negative Control Mating: Y2HGold [pGBKT7-Lam] and Y187 [pGADT7-T].
- Both colonies were placed in a single 1.5 ml centrifuge tube containing 500 µl of 2X YPDA and vortexed to mix.
- Cultures were incubated with shaking at 200 rpm at 30°C overnight.

- From the mated culture (0.5 ml), 100 µl of 1/10, 1/100, and 1/1,000 dilutions were spread on each of the following agar plates. Plates were incubated (colony side facing downward) at 30°C for 3–5 days.
    - SD/–Trp
    - SD/–Leu
    - SD/–Leu/–Trp (=DDO)
    - SD/–Leu/–Trp/X-α-Gal/AbA (=DDO/X/A)
- From DDO plates, 2 mm healthy colonies were picked and restreaked onto fresh DDO plates, and incubated at 30°C for 4 days for possible further use.

2.2.4 Yeast two hybrid mating and screening

For mating, *Saccharomyces cerevisiae* Y2HGold and Y187 strains were used.

In the mating step, a concentrated bait culture was prepared and this bait culture was then combined with a 1 ml Human Fetal Brain cDNA library (Normalized Mate&Plate^{TM} Library, Clontech) aliquot. After 22 hours incubation, the mated culture was plated. We used following procedure;

- One fresh, large (2–3 mm) colony of bait strain (NFIB-α3) was inoculated into 50 ml of SD/–Trp liquid medium.
- The culture was incubated shaking (250–270 rpm) at 30°C until the OD_{600} reaches 0.8 (16–20 hr).
- The culture was centrifuged to pellet the cells (1,000 *g* for 5 min), and supernatant was discarded.
- The pellet was resuspended to a cell density of $2.9x10^8$ cells per ml in SD/–Trp (5 ml). Cells were counted using a hemocytometer.
- A 1 ml aliquot of human fetal brain cDNA library was thawed in a room temperature water bath. 10 µl of aliquot was removed for titering on 100 mm SD/–Leu agar plates.
- 1 ml of our Mate & Plate Library was combined with 5 ml bait culture in a sterile 2 L flask.
- 45 ml of 2xYPDA liquid medium (with 50 µg/ml kanamycin) was added.
- Cells from the library vial was rinsed twice with 1 ml 2xYPDA and added to the 2 L flask.
- Culture was incubated at 30°C for 20–24 hr, slowly shaking (30–50 rpm).

- After 20 hours, a drop of the culture was checked under a phase contrast microscope (40X) to see whether or not zygotes were present, then continued incubation for additional 4 hours.
- The culture was centrifuged to pellet the cells (1,000 *g* for 10 min).
- The 2L flask was rinsed twice with 50 ml 0.5X YPDA (with 50 µg/ml kanamycin), the rinses were combined, and used this to resuspend the pelleted cells.
- The suspension was centrifuged to pellet the cells (1,000 *g* for 10 min) and the supernatant was discarded.
- All pelleteted cells were resuspended in 10 ml of 0.5X YPDA/Kan liquid medium. The total volume of cells + medium was measured.
- To calculate the number of clones screened, 100 µl of 1/10, 1/100, 1/1,000, and 1/10,000 dilutions of the mated culture were spread on each of the following 100 mm agar plates and incubated at 30°C for 3–5 days:
    - SD/–Trp
    - SD/–Leu
    - SD/–Leu/–Trp (DDO)
- The remainder of the culture was plated, 200 µl per 150 mm on DDO/X/A (54 plates) and incubated at 30°C for 3–5 days.
- The number of screened clones (diploids) was calculated by counting the colonies on the DDO plates after 3 days of growth and using the following formula:
    - Number of Screened Clones = cfu/ml of diploids x resuspension volume (ml)
- The mating efficiency was determined as percentage of diploids by first determining the limiting partner of the mated yeast strains and then using the following formula:

$$\frac{\text{No. of cfu/ml of diploids x 100}}{\text{No. of cfu/ml of limiting partner}} = \text{\% Diploids}$$

- We measured the following viability values of the strains (bait or prey) and then the strain (prey or bait) with the lower viability was chosen as limiting partner.
    - No. of cfu/ml on SD/–Leu (viability of the Prey Library)

- No. of cfu/ml on SD/–Trp (viability of Bait)
- No. of cfu/ml on SD/–Leu/–Trp (viability of diploids)

2.2.5 Revealing positive interactions

The QDO/X/A colonies were further analyzed to identify duplicates and to verify positive interactions. To do this, higher stringency screening, yeast colony PCR, restriction digestion, and multiple streaking experiments were performed.

2.2.5.1 Higher stringency screening

All the blue colonies that grew on DDO/X/A were streaked onto higher stringency QDO/A agar plates using a yellow pipette tip. The resulting colonies were then streaked onto QDO/X/A plates.

2.2.5.2 Yeast colony PCR

Yeast Colony PCR was conducted for revealing colonies that contained more than one prey vector. The primers were complementary to upstream (MATCHMAKER 5' AD LD-Insert Screening Amplimer) and downstream (MATCHMAKER 3' AD LD-Insert Screening Amplimer) regions from the prey inserts (referred to as 5-AD-LD-ISA and 3-AD-LD-ISA, respectively).

The following protocol was used for Yeast Colony PCR;

- 100 µl of sterile, ultrapure water was added to a microcentrifuge tube and a small colony was suspended in this tube by vortexing.
- PCR reaction mix was prepared in advance as seen in Table 2.14 and 8 µl of the colony suspension was transferred to the PCR mix.

Table 2.14 Components of yeast colony PCR.

Components	Volume / Amount
dH_2O	10 µl
10X Advantage II PCR Buffer	2.5 µl
50 X dNTP Mix	0.5 µl
10 µM Forward Primer	0.5 µl
10 µM Reverse Primer	0.5 µl
10 X Melting Solution	2.5 µl
50 X Advantage II Polymerase Mix	0.5 µl
Template DNA	8 µl
TOTAL	**25 µl**

:

The following protocol was used for colony PCR;

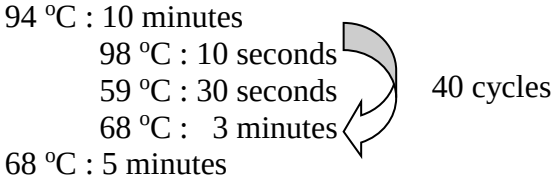


Figure 2.10 Colony PCR cycling protocol

After Colony PCR, 10 µl of PCR product was run on 1% agarose gel stained with 1X electrophoresis dye (PronaSafe) and the colonies that gave more than one PCR product were noted.

2.2.5.3 Restriction digestion of colony PCR products

Once Yeast Colony PCR was completed and potential duplicate clones were determined, PCR products of similar length were subjected to restriction digest. To do this, a frequent cutter enzyme, HaeIII was used. HaeIII is a four-cutter restriction enzyme with the following recognition pattern.

```
5′… GG/CC …3′
3′… CC/GG …5′
```

Figure 2.11 Restriction pattern of HaeIII enzyme

The reactions were set-up as indicated below and incubated for 5 minutes at 37ºC in a water bath.

Table 2.15 Components in the restriction digestion.

Components	Volume / Amount
dH_2O, Nuclease-free	17 µl
10X Fast Digest Green Buffer	2 µl
HaeIII Fast Digest Restriction Enzyme	1 µl
Colony PCR Product	10 µl
TOTAL	**30 µl**

10 µl of the restriction digest was run on 2% agarose gel stained with PronaSafe dye to identify PCR products with similar rstriction digest patterns. When more than one PCR product showed the same restriction pattern, the colonies were deemed to be duplicates and duplicate colonies were eliminated from further analysis.

2.2.5.4 Multiple streaking

Yeast cells may take up more than one version of a plasmid following transformation. This means that a yeast cell can contain a prey vector whose protein product is able interact with the bait and activate reporters, as well as multiple prey vectors whose protein products are noninteracting.

If the DNA of such a yeast clone is tried to be rescued directly via *E.coli* transformation without an intermediate plasmid segregation step, one may select a random *E.coli* colony which contains the wrong plasmid whose protein is not able to interact bait protein. In order to avoid isolation of noninteracting plasmids, one needs to perform segregation steps to increase the chance of selecting positive prey plasmid.

We have streaked blue colonies twice on DDO/X (no Aureobasidin A), each time picking a single blue colony for restreaking. After the first streak, a mixture of blue and white colonies were obtained, indicating segregation of positive interactors (blue) from non-interactors (white). At the end of 2 streaks, most of the colonies appeared in blue color and we continued with them in the next step. The clones none of whose colonies appeared in blue color were eliminated.

We also picked blue colonies from DDO/X plates and streaked them on QDO/A plates twice to make sure that the colonies retained their ability to activate AUR1-C, HIS3 and ADE2 reporters. Only then we proceeded with plasmid rescue from each one of these clones.

2.2.6 Distinguishing genuine interactions

Once the duplicate clones were eliminated by colony PCR, and the non-interacting prey plasmids by library plasmid segregation, we then proceeded with final yeast mating experiments to distinguish genuine interactions from false positives.

2.2.6.1 Final yeast mating

To distinguish genuine interactions from false positives, we have performed a final yeast mating experiment. First, we isolated plasmids, then we transformed them into *E.coli* cells and eventually we performed yeast mating. For this mating step, we have followed the same protocol as we did in the control mating experiments.

Plasmid Isolation from yeast

Before performing final yeast mating experiments, we isolated the library plasmids from colonies isolated after multiple streaking. Since yeast has thick cell walls and their plasmids are inside their spheroplasts, one must apply sequential steps to isolate yeast plasmid DNA. A combination of plasmid isolation protocols was performed. The yeast plasmid isolation procedure is given below;

- One blue colony was picked from QDO/X/A plate and suspended in 1.5 ml DDO medium.
- The suspension was incubated at 30^{o}C overnight shaking at 150 rpm.
- The culture was transfererd to a microcentrifuge tube and centrifuged at 13.500 rpm for 5 minutes.
- The supernatant was poured off and residual medium was removed with a pipette.
- The pellet was resuspended in 50 µl Potassium Phosphate [67 mM KH_2PO_4; pH: 7.5].
- 2 µl of lyticase solution (25 U/µl) diluted in 8 µl TE buffer and added to the tube. The solution was pipetted up and down repeatedly to thoroughly resuspend the pellet.
- The suspension was incubated at 37°C for 60 minutes.
- 20 µl of 10% SDS was added to each tube and vortexed vigorously for 1 min.
- Samples were put through one freeze/thaw cycle (–20°C) and vortexed again before use.
- We proceeded with Roche Plasmid Isolation Protocol described in section 2. starting from suspension step by adding 250 µl suspension buffer.
- At the end, samples were eluted in of 30 µl of elution buffer.

***E.coli* transformation**

E.coli was transformed with isolated yeast plasmids according to the following procedure;

- 20 µl of isolated plasmid and 50 µl competent cells were combined and mixed in a microcentrifuge tube.
- 1 µl of 40% DMSO was added to the tube and mixed.

- The tube was placed on ice bath for 30 minutes.
- A heat shock was applied at 42°C for 90 seconds in a water bath.
- The tube was incubated on an ice bath for 2 minutes.
- 400 µl SOC medium was added to the tube and the tube was incubated with shaking for 45 minutes at 37°C.
- 100 µl of suspension was transferred on an LB plate which contained 100 µg/ml ampicillin and incubated overnight.
- Next day, 3 ml of LB/Amp liquid medium was inoculated with one colony and incubated at 37°C overnight with shaking at 180 rpm.
- The culture was then centrifuged and prey plasmid was isolated according to the miniprep plasmid isolation protocol described in section 2.2.1.7.

Final mating of yeast strains

Since two-hybrid screen can detect false positives, it is important to confirm that the interactions are genuine. If prey activates the Gal4-responsive reporters in the absence of bait, then the interaction is considered to be a false positive. If both bait and prey are required to activate the Gal4-responsive reporters then the interaction is a genuine positive.

The following protocol was used for final yeast mating to distinguish genuine positives from false positives;

- 100 ng of each prey vector was transformed into Y187 Competent Cells using the small scale yeast transformation procedure described in section 2.2.2 and plates were incubated at 30°C until colonies become 3 mm in diameter.
- One fresh bait colony from a SD/-Trp plate was mixed with one Y187 colony transformed with a prey vector in a microcentrifuge tube containing 500 µl 2X YPDA.
- The same procedure was performed by combining prey colonies with a Y2H Gold colony containing empty pGBKT7.
- The microcentrifuge tubes were incubated at 30°C for 24 hours shaking at 150 rpm.
- The mated cultures were diluted 1/10 in 0.9% NaCl solution and 100 µl of diluted mix was spread on QDO/A and DDO plates, incubated for 3 days at 30°C.

At the end of 3 days, prey clones that resulted in colonies when mated with NFIB2 expressing bait clone, and no colony when mated a clone transformed with empty vector were considered to be genuine interactors.

2.2.7 Sequence analysis

Once the genuine positive clones were identified by yeast mating, the prey plasmids isolated from these clones were sequenced to identify these interactors. Sequencing analysis was performed with T7 Sequencing Primer by IonTek, Turkey. The resulting sequences were aligned against GenBank via NCBI Blast Tool (http://blast.ncbi.nlm.nih.gov).

3.RESULTS

3.1 Construction of Bait Plasmid

We cloned mouse Nfib3 cDNA and Nfib3 transcription modulation domain (aa 209-420) as they are cloned in pCHNFI-B2 vector, in this thesis they will referred to as NFIB2 and NFIB2Δ, into pGBKT7 vector by PCR amplifying them from pCHNFI-B2 vector. First, we restriction digested pGBKT7 vector with SalI and XmaI restriction enzymes, then we amplified NFIB2 and NFIB2Δ cDNA fragments from pCHNFI-B2 vector, and eventually, we cloned them to linearized pGBKT7 vector with Clontech InFusion Cloning System.

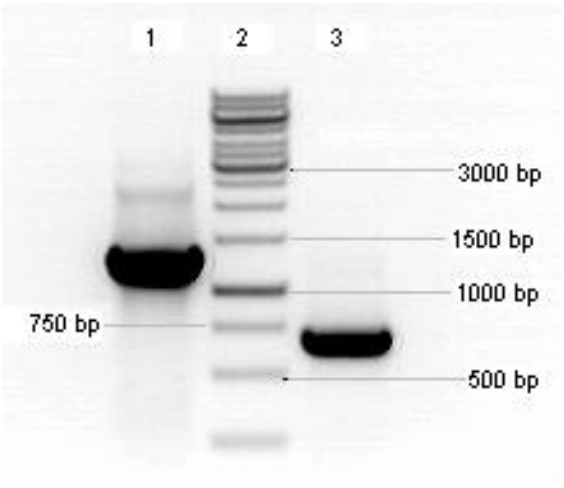


Figure 3.1 Gel image of PCR amplified NFIB2 insert (Lane 1: NFIB2 cDNA PCR-amplified from pCHNFI-B2 Vector, Lane 2: Fermentas GeneRuler^{TM} 1 kb DNA Ladder, Lane 3: NFIB2Δ cDNA PCR-amplified from pCHNFI-B2 Vector).

Once the InFusion cloning reaction was complete, a fraction of the reaction mix was transformed into *E.coli*. Only 3 colonies were obtained for pGBKT7-NFIB2 cloning reaction and 2 colonies for pGBKT7-NFIB2Δ (none for negative control). Plasmid DNA was purified from these colonies and checked for presence of insert by restriction digestion (Figure 3.2).

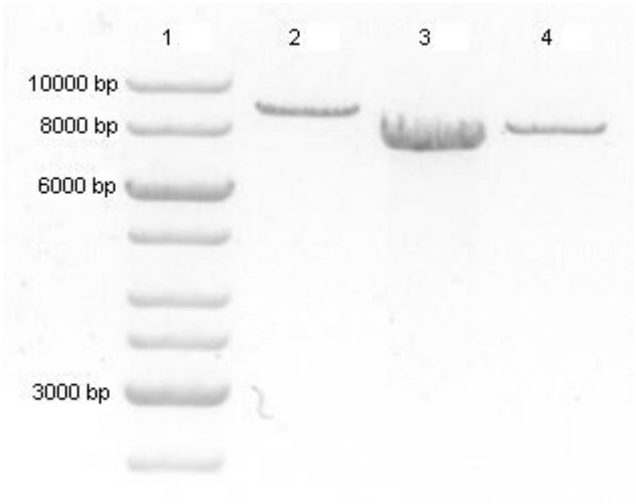


Figure 3.2 Gel image of cloned NFIB2 insert (Lane 1: Fermentas GeneRuler^{TM} 1 kb DNA Ladder, Lane 2: Linearized pGBKT7 with NFIB2 Insert, Lane 3: Linearized pGBKT7, Lane 4: Linearized pGBKT7 with NFIB2Δ Insert).

All 5 colonies were deemed to contain inserts and were subjected to DNA sequencing analysis to check if NFIB2 and NFIB2Δ were cloned into pGBKT7 expression vector correctly and if their sequence were the same as its original sequence in pCHNFI-B2 vector. Seeing that the sequences of cloned fragments corresponded to the expected NFIB2 and NFIB2Δ sequences and were cloned in frame, we proceeded with the control experiments. The sequence data for pGBKT7-NFIB2 and pGBKT7-NFIB2Δ are shown in Appendix A.

3.2 Control Experiments

Before performing the yeast two hybrid screen, we conducted a series of contol experiments. These experiments comprised control for bait expression, control for bait toxicity, control for bait autoactivation, and control for mating.

The bait vector must express GAL4-DNA-BD-NFIB2 fusion protein in order to detect interactions with prey proteins. We verified that Y2HGold colonies transformed with pGBKT7-NFIB2 expressed NFIB2 fusion protein by Western Blot analysis using GAL4-DNA-BD and α-NFI antibodies as shown in Figure 3.3. However, we could not detect a protein product of NFIB2 fusion protein with anti-myc antibody (see section 4 for discussion). Additionally, the truncated version of fusion protein (NFIB2Δ) could not be detected with the α-NFI antibody as this antibody recognizes the N terminal domain.

Several colonies were tested for their ability to express NFIB2 fusion protein and the highest expressing clone (NFIB2-α3) was selected for library screening. The appearent molecular weights of the protein bands, which were determined by their relative mobility in SDS-PAGE with respect to protein molecular weight markers, were in accordance with the expected sizes of the bands as depicted in the Figure below.

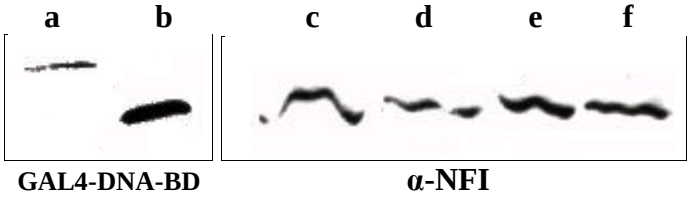


Figure 3.3 Control test for protein expression (**a-b**:Proteins labeled with GAL4-DNA-BD antibody; **c-f**: Proteins labeled with α-NFI antibody; **a,c,d,e,f**: NFIB Protein fused to GAL4 DNA Binding Domain: 63 kDa; **b**:Murine p53 Protein: 57 kDa)

To check whether bait vectors are toxic to the cell we performed the following test. Y2HGold competent cells were transformed with pGBKT7-NFIB2, pGBKT7-NFIB2Δ constructs or the pGBKT7 vector and grown on SD/-Trp plates. After 4 days, the number of resulting colonies were counted and diameters of colonies transformed with different vectors were compared. According to the results as shown in Table 3.1, there was no evidence of toxicity as the colonies of fusion proteins were similar to colonies of no fusion protein in number and in diameter.

Table 3.1 Results of toxicity test

Vector	Dilution	Medium	Number of Colonies	Diameter	Efficiency cfu/ug
pGBKT7	1/10	SD/-Trp	69	2mm	$6{,}9\times10^{4}$
pGBKT7-NFIB2	1/10	SD/-Trp	77	2mm	$7{,}7\times10^{4}$
pGBKT7- NFIB2Δ	1/10	SD/-Trp	50	2mm	$5{,}0\times10^{4}$
pGBKT7	1/100	SD/-Trp	10	2mm	$1{,}0\times10^{5}$
pGBKT7-NFIB2	1/100	SD/-Trp	12	2mm	$1{,}2\times10^{5}$
pGBKT7- NFIB2Δ	1/100	SD/-Trp	14	2mm	$1{,}4\times10^{5}$

In the autoactivation test, we tested if bait protein autonomously activates the reporter genes in Y2HGold in the absence of a prey protein. According to the results given in Table 3.2., NFIB2 did not autoactivate, however, there was a significant autoactivation by the NFIB2Δ fusion protein as SD/-Trp/X-α-Gal/AbA plate gave

blue colonies. Therefore, following control tests, we performed mating experiments only with NFIB2 fusion protein.

Table 3.2 Results of autoactivation test

Vector	Medium	Color of Cells
pGBKT7-NFIB2	SD/-Trp	White cells
	SD/-Trp/X	A mix of blue and pale blue cells
	SD/-Trp/X/A	No cell
pGBKT7-NFIB2Δ	SD/-Trp	White cells
	SD/-Trp/X	Blue cells
	SD/-Trp/X/A	Blue cells

Finally, a mating control was conducted to optimize the procedures of the two-hybrid screen. For positive mating, pGBKT7-53 that encodes the Gal4 DNA-BD fused to murine p53 and pGADT7-T that encodes the Gal4 AD fused with SV40 large T-antigen were mated. We also performed a negative control using pGBKT7-Lam (which encodes the Gal4 BD fused with lamin) and pGADT7-T. The following results were obtained from mating control:

Table 3.3 Results of mating control

Vector 1	Vector 2	Dilution	Medium	Colony Number
pGBKT7-53	pGADT7-T	1/100	SD/-Leu /-Trp/X-α-gal-AbA	10
pGBKT7-53	pGADT7-T	1/100	SD/-Leu /-Trp	18
pGBKT7-53	pGADT7-T	1/100	SD/-Leu	>1000
pGBKT7-53	pGADT7-T	1/100	SD/-Trp	>1000
pGBKT7-53	pGADT7-T	1/10	SD/-Leu /-Trp/X-α-gal-AbA	96
pGBKT7-53	pGADT7-T	1/10	SD/-Leu /-Trp	176
pGBKT7-53	pGADT7-T	1/10	SD/-Leu	>2000
pGBKT7-53	pGADT7-T	1/10	SD/-Trp	>2000

According to the results, the colonies were mated properly.

3.3 Yeast Two Hybrid Mating and Screening

To screen for NFIB binding proteins, a concentrated bait culture of NFIB2-α3 was prepared and the library aliquot was thawed.

After number of cells in bait and prey cultures were counted, cultures were combined in a sterile 2L flask. At this stage, the concentrations of bait culture and prey aliquot were 2.9×10^8 cell/ml and 1.39×10^9 cell/ml, respectively.

The process of mating was first checked at 20-hours. Since only small number of diploids were seen under microscope, we waited an additional 2 hours and then harvested the cells. Finally, we resuspended cells in 10 ml 0.5X YPDA/Kan liquid medium resulting in a total suspension volume of 13.5 ml. This suspension was spread on 54 × 150 mm DDO agar plates and incubated at 30°C for 3 days.

3.3.1 Calculating mating efficiency

A mating efficiency between 2–5% is reported to indicate succesful mating. In order to calculate the mating efficiency, one needs to first determine the limiting partner among prey and bait cultures. The limiting partner is the one its *number of cfu/ml* value is less than the other partner. The formula to calculate *number of cfu/ml* is given below;

Number of cfu/ml = [Number of cfu × 1000 µl/ml] / [Volume Plated (µl) × Dilution factor]

According to our results, the limiting parner is the prey library. Since less diluted plates resulted in too many colonies to count, we only focused on 1/10,000 dilution plates. Our results were as follows;

1/10,000 SD/-Leu/-Trp → 24 (for diploids)
1/10,000 SD/-Leu → 224 (for prey)
1/10,000 SD/-Trp → 2000 (for bait)

Applying the formula above, we found the following *Number of cfu/ml* values;

Number of cfu/ml for SD/-Leu/-Trp → 2.4×10^6
Number of cfu/ml for SD/-Leu → 2.24×10^7
Number of cfu/ml for SD/-Trp → 2×10^8

Since mating efficiency is expressed as the ratio of *number of dipoid cells* to the *number of cells of limiting partner*, our mating efficiency was as follows;

$$\text{Mating Efficiency} = [2.4 \times 10^6] / [2.24 \times 10^7] = 10.7\%$$

3.3.2 Counting the number of clones screened

The number of screened clones (diploids) was calculated by counting the colonies from the DDO plates (24 colonies on 1/10,000 dilution plates) after 4 days and applying following formula;

Number of Screened Clones = cfu/ml of diploids × resuspension volume (ml)

According to this formula, the number of clones we screened was as follows:

$$\text{Number of Screened Clones} = 2.4 \times 10^6 \times 13.5 = 32.4 \text{ million}$$

3.4 Confirmation of positive interactions

To confirm positive interactions, we performed higher stringency streaking, yeast colony PCR, restriction digestion, and multiple streaking experiments.

3.4.1 Plating on higher stringency medium

470 blue colonies from 150 mm plates were streaked on higher stringency agar plates, QDO/A. At the end of 3 days, 258 grew and gave blue colonies. When these were streaked on QDO/X/A plates, 57 blue colonies were obtained. These 57 colonies were further analyzed by yeast colony PCR.

3.4.2 Yeast colony PCR

Using Matchmaker™ AD LD-Insert Screening Amplimers, we performed colony PCR to screen for colonies that contained more than one prey vector. PCR products were analyzed by agarose gel electrophoresis (Figure 3.4).

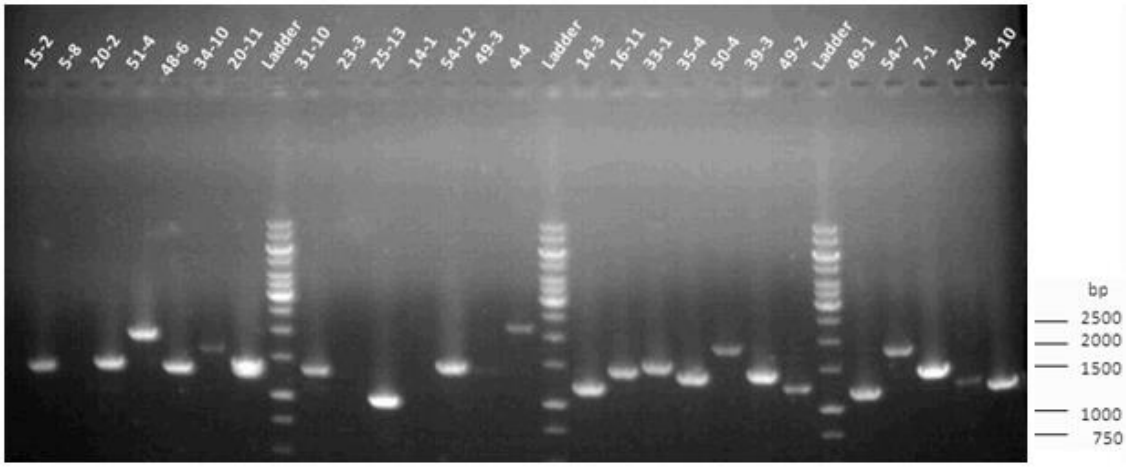


Figure 3.4a Gel image of yeast colony PCR (10 µl of Colony PCR product was run on 1% agarose gel)

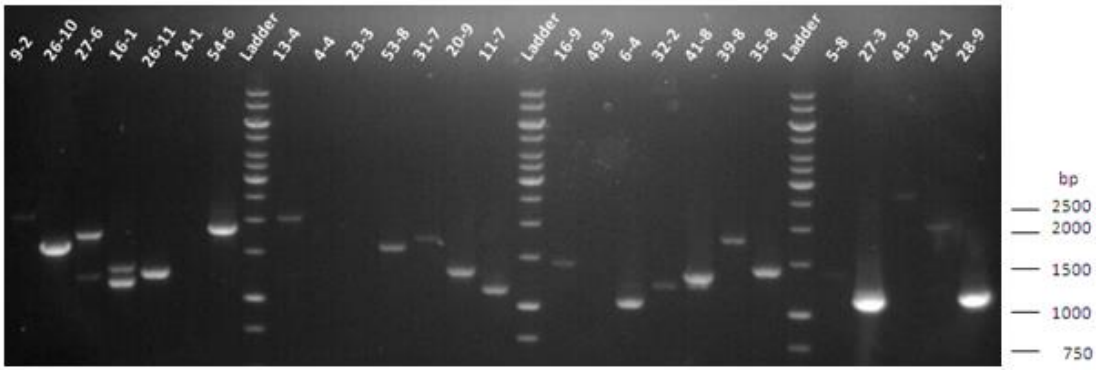


Figure 3.4b Gel image of yeast colony PCR (10 µl of Colony PCR product was run on 1% agarose gel)

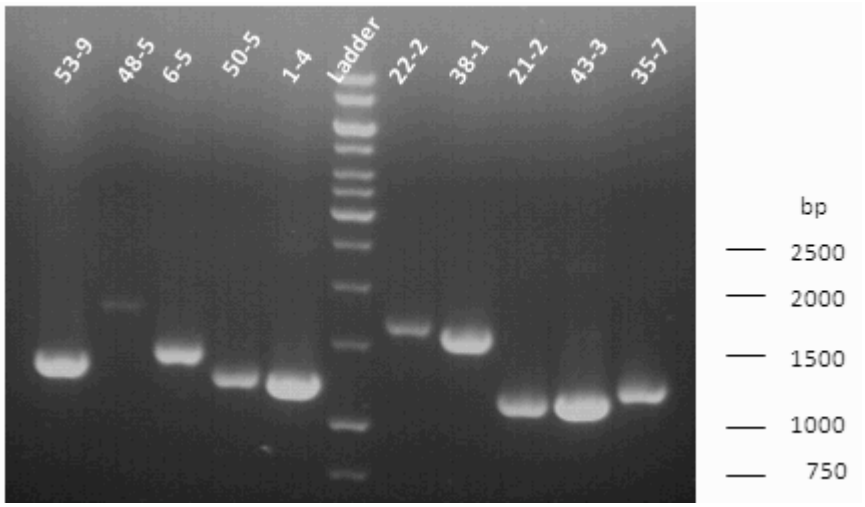


Figure 3.4c Gel image of yeast colony PCR (10 µl of Colony PCR product was run on 1% agarose gel)

As seen in the images, most of the bands are in the 1000 bp-2000 bp range.

3.4.3 Restriction digestion of colony PCR products

Restriction digestion analysis of yeast colony PCR products was performed with HaeIII. Following restriction reaction, the digested samples were analyzed by agarose gel electrophoresis as shown in Figure 3.5.

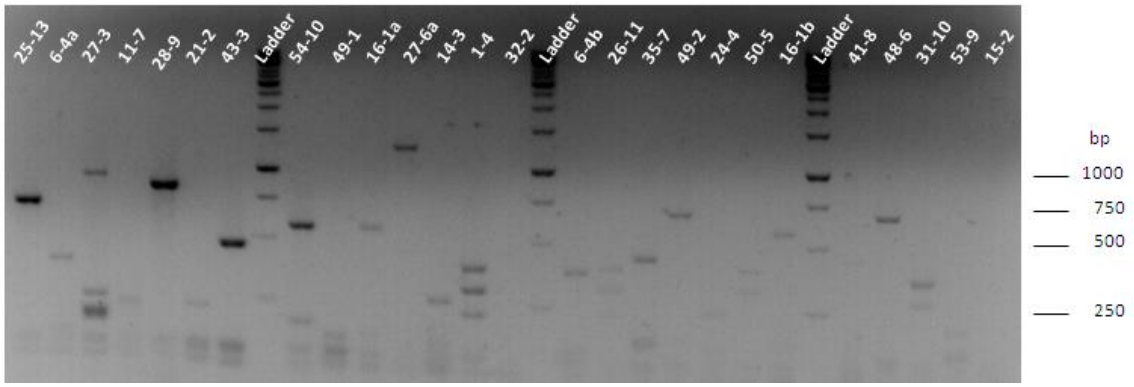


Figure 3.5a Gel image of restricted fragments with HaeIII (10 µl of restriction product was run on 2% agarose gel)

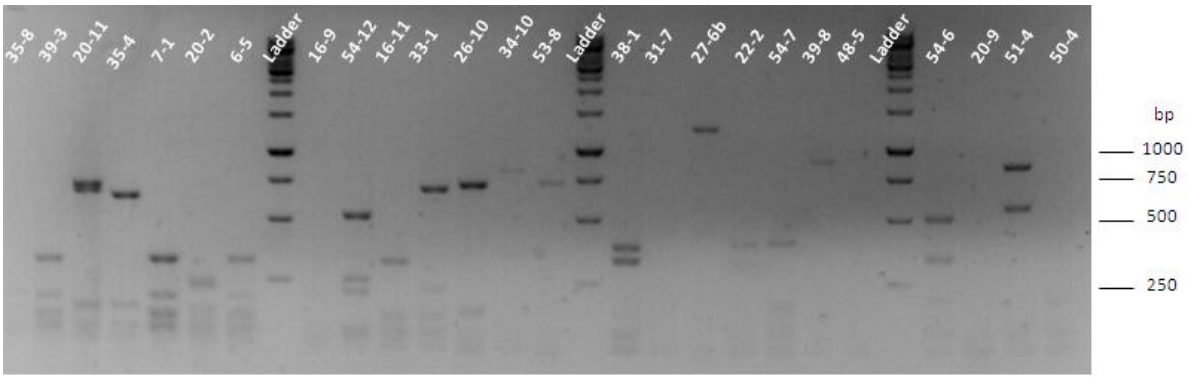


Figure 3.5b Gel image of restricted fragments with HaeIII (10 µl of restriction product was run on 2% agarose gel)

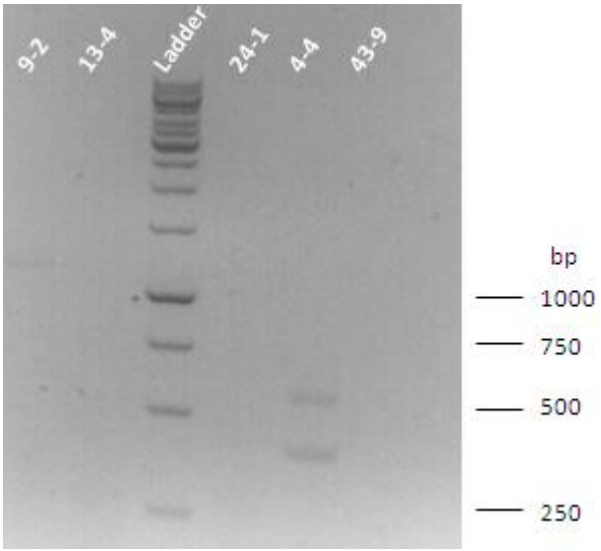


Figure 3.5c Gel image of restricted fragments with HaeIII (10 µl of restriction product was run on 2% agarose gel)

Based on the information obtained from restriction digestion as seen in the Figure 3.5a and 3.5b and 3.5c, some of clones were thought to be the identical in length and these were then run on agarose gel again. The ultimate comparison of these clones are shown in Figure 3.6a and 3.6b.

5 of the total number of colonies (39-3, 7-1, 6-5 and 1-4, 26-11) were determined to have had the same digestion pattern, and 3 of these clones were eliminated.

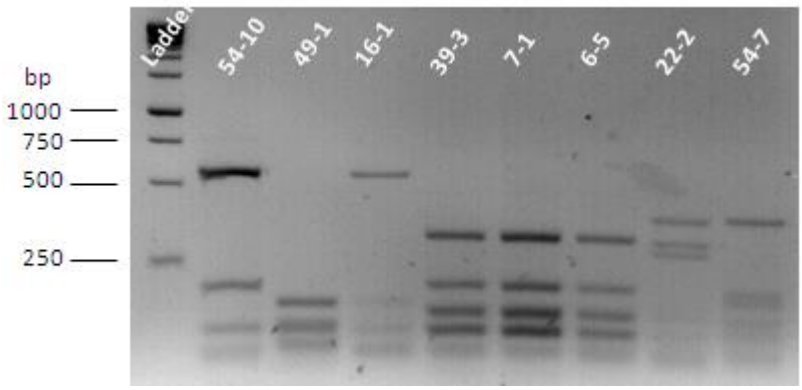


Figure 3.6a Gel Image of restriction fragments of similar length (10 µl of restriction product was run on 2% agarose gel stained with PronaSafe dye)

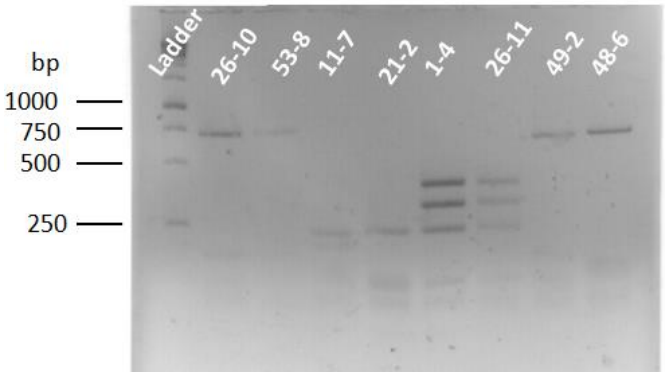


Figure 3.6b Gel image of restriction fragments of similar length (10 µl of restriction product was run on 2% agarose gel stained with PronaSafe dye)

3.4.4 Multiple streaking

Multiple streaks were performed to increase the chance of rescuing the positive prey plasmid. For this, blue colonies from QDO/X/A plates were restreaked on DDO/X, each time picking a single blue colony for restreaking. At the end of 2 streaks, the total number of colonies decreased to 49 as 5 colonies were white only. Next, these clones were tested for their ability to activate AUR1-C, HIS3 and ADE2 reporters. To do this, we picked blue colonies from DDO/X plates and streaked them on QDO/A plates. The number of surviving clones decreased to 40. Then, we proceeded with plasmid rescue from each one of these clones.

3.5 Distinguishing Genuine Interactions

To distinguish genuine interactions from false positives, we performed yeast mating experiments.

3.5.1 Yeast mating

In order to reveal genuine interactions between bait and potential prey canditates, we performed final yeast mating experiments. For this, we used bait colonies of NFIB2-α3 and prey colonies of Y187 into which we previously transformed prey vectors.

The number of prey clones that we investigated in this final mating procedure was 40 and 24 of these clones which were mated with NFIB2-α3 gave colonies on QDO/A plates but no colonies on QDO/A plates when mated with empty vector (outlined in Table 3.4, underlined numbers correspond to clones which were considered as positive). The number of colonies were greater than 1000 on all DDO plates. We then sent the plasmid DNAs of these positive clones to sequencing.

Table 3.4 Plate results of final yeast mating

Colony Number	Control Plate ↔ NFIB2-α3 Plate	Colony Number	Control Plate ↔ NFIB2-α3 Plate
34-10	0 ↔ >100	24-4	78 ↔ 30
16-11	0 ↔ 4	20-11	134 ↔ >400
1-4	0 ↔ >100	48-6	28 ↔ 162
54-7	0 ↔ >100	9-2	>100 ↔ >100
49-2	0 ↔ >100	7-1	>100 ↔ >100
21-2	0 ↔ 26	35-8	0 ↔ 13
28-9	2 ↔ 26	54-6	0 ↔ 24
31-10	8 ↔ 41	26-10	>100↔ >200
31-7	>1000↔ >1000	53-8	>1000 ↔ >1000
27-3	0 ↔ 0	22-2	0 ↔ >100
53-9	0 ↔ 0	48-5	0 ↔ >100
25-13	0 ↔ 36	51-4	0 ↔ >100
50-5	0 ↔ 52	13-4	17 ↔ >200
38-1	0 ↔ 38	24-1	16 ↔ >100
20-9	0 ↔ >100	49-3	0 ↔ >100
23-3	0 ↔ >100	5-8	0 ↔ >100
16-9	0 ↔ >100	16-1	0 ↔ >100
14-1	0 ↔ 28	14-3	0 ↔ >100
4-4	1 ↔ 218	11-7	>1000 ↔ >100
43-3	>1000 ↔ 179	32-2	0 ↔ >100

3.6 Sequence Analysis

Out of 40 prey clones identified with yeast two hybrid screening, 24 were sent to sequencing. 4 clones were detected to code for the same protein, 4 clones corresponded to noncoding genomic sequences, 3 clones were not in frame with Gal4

activation domain and one clone corresponded to 3'UTR of a protein and they were eliminated. Consequently, 12 were found to code for a full or a partial protein. These proteins are listed below:

Homo sapiens LIM domain only 4 (LMO4), mRNA

LMO4 gene encodes a cysteine-rich protein that contains two LIM domains but lacks a DNA-binding homeodomain. It is a transcription factor and it shows its transcription modulation activity by interacting selectively and non-covalently with a specific DNA sequence, with a transcription factor, with an enhancer, or with zinc (Zn) ions.

LMO4 plays essential roles in the developing brain. The loss of LMO4 in the cortex results in neurogenesis defects whereas its expression facilitates differentiation of cortical neurons (Asprer et al., 2011). Its function has been shown in neural tube closure (Hahm et al.,2004; Lee et al., 2005) and development of the neural tube in mice (Tse et al., 2004). It was proposed to have role in the diversity of motor cortex projection neuron subpopulations (Cederquist et al., 2013) and in the balanced generation of inhibitory and excitatory neurons in the ventral spinal cord (Joshi et al., 2009).

The effects of LMO4 in neurodegenerative diseases were also investigated. Its high level of expression in the entorhinal cortex (EC) and in the CA1 hippocampal region was seen in healthy brains whereas a consistent decrease in the expression of LMO4 was observed in AD brains (Leuba et al., 2004). In protecting neurons from ischemic injury after stroke requires LMO4 (Schock et al., 2008).

LMO4 was found to interact with several proteins such as CREB in calcium-dependent gene expression (Kashani et al.2006), nuclear LIM interactor (NLI) in mediating nuclear retention of LMO proteins in the early embryo (Kenny et al., 1998), basic Helix-Loop-Helix protein (HEN1) in modulating transcriptional activity of HEN1 in neuronal cells (Manetopoulos et al., 2003), peroxisome proliferator-activated receptor gamma (PPARγ) in PPARgamma signaling to protect cortical neurons from ischemic injury (Schock et al., 2008) and Neogenin in repulsive axon guidance in human neuronal cells and in rat primary neurons (Schaffar et al., 2008).

The expression of LMO4 was detected in many tissues throughout human body but particularly high expression was seen in fetal brain and prefrontal cortex (Url-1). The

localization of LMO4 at sub-cellular level was detected primarily in the nucleus with a substantial amount also detected in the cytoplasm (HPRD: 09122).

Homo sapiens prosaposin (PSAP), transcript variant 3, mRNA

This gene encodes a highly conserved glycoprotein which is a precursor for 4 cleavage products: saposins A, B, C, and D. They localize primarily to the lysosomal compartment where they facilitate the catabolism of glycosphingolipids with short oligosaccharide groups. It exists both as a secretory protein and as an integral membrane protein. Its greatest expression in the brain was seen in hypothalamus, thalamus and prefrontal cortex.

PSAP is suggested to be an essential neurotrophic factor and vital for the survival of cortical neurons as its deficiency resulted in massive neuronal loss (Sikora et al., 2007). $PSAP^{-/-}$ mice was shown to have a short life span (30 days) and central nervous system (CNS) neuronal degeneration (Chu et al., 2005). One of four glycoproteins (saposin C) encoded by PSAP was suggested to play a role in axonal membrane homeostasis, as disruption of saposin C resulted in neurodegeneration in lysosomal storage disease (Yoneshige et al., 2010). It was proposed that neural progenitor cells secrete PSAP to defend neighboring mature neurons from apoptosis upon injury (Li et al., 2010).

By performing yeast two-hybrid screen using human fetal brain cDNA, PSAP was found to ineract with TFPI-2 (tissue factor pathway inhibitor-2) and this interaction inhibits PSAP, leading to the inhibited abilities of invasion, migration and even metastasis of malignant tumour cells (Xu et al., 2012). Another yeast two-hybrid screen revealed that PSAP interacts with Rhox5 and this interaction was hypotesized to regulate the development of male reproductive organs dynamically (Guo et al., 2007).

Mutations in this gene have been associated with Gaucher disease, Tay-Sachs disease, and metachromatic leukodystrophy.

Homo sapiens peptidylprolyl cis/trans isomerase, NIMA-interacting 1 (PIN1), transcript variant 1, mRNA

PIN1 gene encodes one of the PPIases (Peptidyl-prolyl cis/trans isomerases) that specifically binds to phosphorylated ser/thr-pro motifs to catalytically regulate the

post-phosphorylation conformation of its substrates via cis/trans isomerization of peptidyl-prolyl peptide bonds.

Pin1 was shown to be highly expressed in NPCs at the late developmental stage and has a major impact on their neuronal differentiation (Nakamura et al., 2010). In differentiating neurons an interaction between BNIP-H and Pin1 was identified which may provide a post-translational regulation on the cellular activity of BNIP-H during neuronal differentiation (Buschdorf et al., 2008). β-catenin was identified as an interactor of Pin1 in neural progenitor cells and upon interaction, Pin1 stabilizes β-catenin by inducing conformational changes after phosphorylation (Nakamura et al., 2010).

Pin1 was also shown to be involved in neurodegenerative disorders. Pin1 was proposed to play a role in Parkinson disease (PD) by mediating Lewy body formations which is an indication of PD (Ryo et al., 2006). PIN1 was shown to interact with tau in Alzheimer's disease (AD) brain and this interaction is thought to restore the ability of tau to bind microtubules and promote microtubule stability (Dakson et al., 2011). Pin1 dysregulation has been proposed to affect the progression of neurodegenerative pathology as Pin1 inhibition was shown to up-regulate cyclin D1 and caspase 3 producing apoptosis in cultured rat hippocampal neurons (Atabay and Karabay, 2012).

While its expression was detected in liver, kidney, blood, heart, expression of PIN1 in the brain is higher than that of in other organs (Url-2). At the sub-cellular level, PIN1 protein is mainly localized in nucleus as well as the cytoplasm (HPRD:03031).

Homo sapiens microtubule-associated protein 1B (MAP1B), mRNA

MAP1B encodes a protein that belongs to the microtubule-associated protein family. The proteins of this family are thought to be involved in microtubule assembly, which is an essential step in neurogenesis.

MAP1B was found to interact with metabotropic glutamate receptor 4 (mGluR4), a G-protein-coupled receptor that mediates inhibition of neurotransmitter release. Since MAP1B and mGluR4 colocalize at excitatory synapses in cultured hippocampal neurons, it was proposed that MAP1B might be implicated in the synaptic trafficking and/or regulation of mGluR4 (Moritz et al.,2009). By two-hybrid and pull-down assays, α-, β-, and βIII-tubulins were found to interact with MAP1B

1-126 and since overexpression of MAP1B 1-126 induced both neurite extension and neuronal death it was proposed that MAP1B 1-126 could be involved in neuronal degeneration (Gomi and Uchida, 2012). A possible interaction between MAP1B and tubulin tyrosine ligase was shown and it was suggested that this interaction may be important for general processes involved in nervous system development such as axonal guidance and neuronal migration (Utreras et al., 2008). Fujiwara et al. found that neuronal Hu proteins directly interact with the light chain of microtubule-associated proteins MAP1B (LC1) and suggested that Hu proteins are involved in microtubule-mediated regulation of mRNA expression within neuronal processes (Fujiwara et al.,2011).

At the sub-cellular level, it is primarly localized at cytoplasm and cytoskeleton and alternatively in nucleus and on plasma membrane and its expression is high in cerebellum, cerebrum, fetus and neuron during development (HPRD: 01139).

Homo sapiens ataxin 3 (ATXN3), transcript variant o, mRNA

ATXN3 has deubiquitinase activity and appears to be a component of the ubiquitin proteasome system. It may also have roles in transcriptional regulation and neuroprotection (reviewed by Haacke et al., 2006).

The protein encoded by ATXN3 contains $(CAG)_n$ repeats in the coding region, and the expansion of these repeats from the normal 13-36 to 68-79 is one of the causes of Machado-Joseph disease（also known as spinocerebellar ataxia-3), an autosomal dominant neurologic disorder (Bettencourt e al., 2012).

By performing yeast two hybrid and GST pull-down assays, it was demonstrated that ATXN3 is able to establish a direct interaction with NEDD8, through its Josephin domain (JD) and it was proposed that this interaction may add a new perspective to the possible physiological role of ATXN3 modulating the activity of the ubiquitin–proteasome system (Ferro et al., 2007). Other identified interactors of ATXN3 are proteasome-binding factors Rad23 and valosin-containing protein (VCP/p97) (Doss-Pepe et al., 2003), HHR23A and HHR23B (Wang et al., 2000).

The expression of ATXN3 in CNS is mainly in caudate nucleus, cerebellum, frontal cortex, pons. While in neurons, ATXN3 was shown to predominantly localize in cytoplasm (Paulson et al.,1997), ATXN3 becomes concentrated in the nucleus of neurons during disease (Bettencourt and Lima, 2011).

SGIP1 SH3-domain GRB2-like (endophilin) interacting protein 1 (SGIP1): SGIP1 is an endocytic protein that affects signaling by receptors in neuronal systems involved in energy homeostasis via its interaction with endophilins (Trevaskis et al., 2005; Uezu et al., 2007). The processes that SGIP1 is involved are endocytosis, positive regulation of energy homeostasis, positive regulation of feeding behavior, positive regulation of receptor-mediated endocytosis and response to dietary excess. It interacts with SH3 domains (Src homology 3) of proteins. Trevaskis et al. (2005) showed that SGIP1 interacts with endophilin-3 and endophilin-1 via the SGIP1 central proline-rich region. Other interactors of SGIP1 are AMPH, Eps15, ITSN1, REPS1. Its expression was detected in many tissues throughout human body such as adipose, blood, breast, colon, kidney, lung, ovary, prostate, skeletal muscle. However, its expression was also seen in brain regions such as cerebellum, frontal cortex and frontal lobe. Trevaskis et al. (2005) performed RT-PCR of P. obesus tissues and detected strong expression in brain, particularly in cerebellum, cortex, and midbrain. At sub-cellular level, SGIP1 protein is localized in the plasma membrane, clathrin-coated pits and in the cytoplasm.

Homo sapiens 3-hydroxy-3-methylglutaryl-CoA synthase 1 (soluble) (HMGCS1), transcript variant 1, mRNA

HMGCS1 is an enzyme that condenses acetyl-CoA with acetoacetyl-CoA to form HMG-CoA, which is the substrate for HMG-CoA reductase. The processes in which HMGCS1 involves are cellular lipid metabolic process, cellular response to cholesterol and hormone stimulus, response to acid and drug, liver and brain development. The expression of this protein in fetal brain is high. At the sub-cellular level, HMGCS1 protein is localized in the cytoplasm.

Homo sapiens nitrogen permease regulator-like 2 (S. cerevisiae) (NPRL2), mRNA

NPRL2 suppresses Src-dependent tyrosine phosphorylation and activation of PDPK1 and its downstream signaling. Its main functions are negative regulation of kinase activity and protein phosphorylation. It is also known as tumor suppressor candidate 4. The expression of a NPRL2 transcript was found to be most abundant in skeletal muscle, followed by brain, liver, and pancreas. At the sub-cellular level, it is primarly localized in the cytoplasm.

Homo sapiens complement component 1, q subcomponent, A chain (C1QA), mRNA

C1QA encodes a major constituent of the human complement subcomponent C1q. C1q associates with C1r and C1s in order to yield the first component of the serum complement system. Deficiency of C1q has been associated with lupus erythematosus and glomerulonephritis. The processes in which C1QA is involved are cell-cell signaling, complement activation, innate immune response. Stevens et al. (2007) found that expression of mouse C1q by postnatal neurons is induced by immature astrocytes and C1q protein localized to synapses throughout the postnatal central nervous system (CNS) and retina. C1QA is an extracellularly localized protein.

Homo sapiens MAD2 mitotic arrest deficient-like 2 (yeast) (MAD2L2), transcript variant 1, mRNA

MAD2L2 is a component of the mitotic spindle assembly checkpoint that prevents the onset of anaphase until all chromosomes are properly aligned at the metaphase plate. In *S. cerevisiae*, members of the MAD (mitotic arrest deficient) encode proteins that play a role in the mitotic spindle checkpoint, its defects have been implicated in the aneuploidy observed in human cancer cells. The processes in which MAD2L2 is involved are DNA repair, mitotic spindle assembly checkpoint, actin filament organization. At the sub-cellular level, it is primarily localized in the nucleus and alternatively in the cytoplasm.

Homo sapiens syndecan binding protein (syntenin) (SDCBP), transcript variant 1, mRNA

SDCBP was initially identified as a molecule linking syndecan-mediated signaling to the cytoskeleton. The syntenin protein contains tandemly repeated PDZ domains that bind the cytoplasmic, C-terminal domains of a variety of transmembrane proteins. This protein may also affect cytoskeletal-membrane organization, cell adhesion, protein trafficking, and the activation of transcription factors. The syntenin-1 protein associates with Neurofascin, which stimulates neurite outgrowth in chicken tectal neurons (PMID: 10662501, PMID: 19356150). This protein is primarily localized to membrane-associated adherens junctions and focal adhesions but is also found at the endoplasmic reticulum and nucleus.

Homo sapiens SEC31 homolog A (S. cerevisiae) (SEC31A), transcript variant 6, mRNA

SEC31A protein is similar to yeast Sec31 protein, a component of the COPII protein complex that is responsible for vesicle budding from endoplasmic reticulum (ER). The main processes in which MAD2L2 is involved are ER to Golgi vesicle-mediated transport, cellular membrane organization, COPII vesicle coating. This protein is primarily localized to the endoplasmic reticulum as well as the cytoplasm, golgi apparatus and cucleus.

4. DISCUSSION

4.1 Potential NFIB Interacting Proteins

The yeast two hybrid stystem revealed 12 genuine interactions. While most of them seem to less likely to interact with NFIB, 5 of them (ATXN3, MAP1B, PIN1, PSAP, LMO4) are strong candidates to be potential NFIB interactors.

LMO4 modulates trancription by binding other transcription factors. Its expression in fetal brain is high and it plays roles in spinal cord and in intermediate neuron differentiation. At sub-cellular level, it localizes in nucleus. So, it seems to be an important canditate in terms of interacting with NFIB.

PIN1 protein codes for one of PPIases (Peptidylprolyl cis-trans isomerases), and regulates the conformation of substrates following phosphorylation. Besides its functions in cell growth, immune and stress response, pluripotent germ cell maintenance and induction, it also plays role in neuron differentation and maintenance. It is also known that this protein is involved in Alzheimer and neurodegenerative disorders. That it is localized in nucleus increases the its possibility to interact with NFIB.

MAP1B protein is primarly localized in cytoplasm, not in nucleus. However since it is involved in microtubule formation, a main process in neurogenesis, it draws attention. Additionally, it is important that this protein was shown to play role in neural development through gene deletion studies and that its expression is high in cerebellum, cerebrum during development.

ATXN3 protein is thought to play role in proteasome system due to its deubiquitinase activity. Interestingly, it is involved in Machado-Joseph disease, an autosomal dominant neurologic disorder. Additionally, it plays role in transcriptional regulation and neuroprotection. It has strong expression in central nervous system and is localized in nucleus. All of this data shows that it may be a real interactor of NFIB.

PSAP is a protein which is involved in processes such as blood coagulation and lipid metabolism. It can function both as a secretary protein and as an integral protein. Altough it does not seem to be involved in processes associated with NFIB, it can show neurotrophic activity, and it is involved in maintenance of neurons and because of these, it may interact with NFIB in nucleus.

SGIP1 plays roles in the endocytosis related processes. It generally bind to membrane associated proteins via its SH3 domain. Altough its expression are seen in cerebellum, cortex, and midbrain, it is localized in membrane and in clathrin coated vesicles. Since NFIB protein is localized in nucleus, the probability that NFIB and SGIP1 interact wih each other is low.

HMGCS1 is basically involved in lipid metabolism, cholesterol and hormonal stimulus. At sub-cellular level, it is located in cytoplasm. Altough it was not proven experimentally, the algorithms in electronic databases reveal that it may play a role in brain development.

NPRL2 is a protein which is involved in protein phosphorylation and regulation of kinase activity. Since it is primarly expressed in skeletal muscle and it was not detected to play a role in neurological processes, it is not expected for this protein to interact with NFIB.

C1QA protein codes for some members of serum complement system which plays role in autoimmune reaction development in humans. It has functions in cell signaling, in activation of complement system, and in immun response. Since C1QA protein plays roles in postnatal neurons, and is subcellularly localized in cytoplasm, not in nucleus, it seems to be a low possibility that this protein interacts with NFIB in developing brain.

MAD2L2 protein plays role in controlling that chromosomes are aligned correctly before anaphase stage in cell cycle as a component of mitotic spindle assembly checkpoint. It is primarly localized in nucleus but no relation of this protein with neurological processes was detected so there is no clue that MAD2L2 may interact with NFIB.

SDCBP protein binds to C-terminal domains of transmembran proteins and is involved in syndecan mediated signaling to cytoskeleton. Beside its role in protein trafficking, cell adhesion, cytoskeletal-membrane organization processes, it is

interesting that this protein is associated with Neurofascin protein which stimulates neurite outgrowth.

SEC31A protein is involved in processes such as transport between endoplasmic reticulum (ER) and Golgi apparatus, vesicle coating and membrane organization. Since it is primarily localized in ER and it was not detected to be involved in neurological processes, it is unlikely that SEC31A protein interacts wih NFIB.

4.2 Technical Issues

4.2.1 Final mating procedure

To prove genuine interactions of prey and bait proteins, we tried to co-transform bait and prey vectors first into competent Y2HGold yeast cells. However, after growing transformed cells on QDO/A medium for 3 days, we had difficulty achieving sufficient transformation efficieny. The number of colonies on DDO plates were very low, and the number of colonies on QDO/A plates were zero indicating that a significant amount yeast cells did not take in the two different plasmids, hindering detection of interactions.

The probability of transforming the same yeast cell with multiple vectors is very low. Therefore, Y2HGold cells that expressed bait protein (NFIB2-α3) were transformed with prey expressing vectors (sequential transformation). While these experiments performed better than simultaneous transformation, the efficiency was still low.

Finally genuine interactors were determined by mating experiments as described for control mating procedures. This “final mating experiments” circumvented the problem and we obtained sufficient number of colonies on QDO/A plates and distinguished genuine interactions.

4.2.2 Autoactivation control

In the autoactivation experiment, we did not see any autoactivation with the plasmid containing NFIB2 cDNA. However, the truncated version of NFIB2 spanning the transcription modulation domain (NFIB2Δ) autoactivated reporters. This construct was then discarded.

It appears that NFIB activation domain (209-420 aa) has the capability to auto-activate these reporters in yeast and that the DNA-Binding Domain of NFIB prevents complete protein from auto-activating reporters.

4.2.3 Control for protein expression of bait vector

Additionally, eventhough both pGBKT7-53 and pGBKT7-NFIB2 constructs contain the same sequence specific to anti-c-myc antibody, we detected pGBKT7-53 cDNA expression with anti-c-myc antibody, while we could not detect the expression of NFIB2 or NFIB2Δ. It is possible that GAL4-DNA-BD- NFIB2 fusion protein folds in such a way that the myc epitope is not exposed. The fact that the NFI antibody detected a protein of expected size and anti GAL4-DBD antibody also detected the same protein indicated that the NFIB2 fusion protein was expressed as expected.

When we tested protein expression of NFIB2Δ fusion protein with α-NFI antibody, we could not detect a band for this protein. The reason for this is that the NFI antibody (H-300) is a rabbit polyclonal IgG its epitope corresponds to amino acids 1-300 mapping at the N-terminus of NFI proteins of human origin. However, the NFIB2Δ fusion protein comprised only the C-terminus of NFIB2 protein.

4.2.4 Yeast two hybrid technique

Y2H is a very powerful technique that is extensively used in protein-protein interaction studies. It is superior to bacterial expression systems as it embodies an *in vivo* technique using the yeast host cell.

In this thesis project, although we have identified novel interaction partners of NFIB, we could not detect the proteins that were previously found to interact with NFIB. Additionally, NFI proteins are expected to interact with each other as they need to form dimers to be functional.

One reason for this false negative interactions may be folding problems. We hypothesized that the presence of Gal4-DNA-BD in fusion protein may have negatively affected correct folding of NFIB. Another possibility is that Y2H screen could not detect transient interactions that normally occur in NFIB's known interactions with other proteins. Lastly, the proteins may not have been cloned in correct reading frame. While in conventional Y2H screens, only one out of six fused cDNAs is in the correct reading frame (Criekinge and Beyaert, 1999), in this screen,

since we used a directional library, one out of three fused cDNAs are expected to be in the correct reading frame.

5. CONCLUSION AND FUTURE DIRECTIONS

In this study, we have isolated protential NFIB binding partners from Human Fetal Brain Library. We determined the identity of interacting proteins by sequencing their cDNAs and recovering the genes coding from these proteins by searching on the NCBI database. However, Yeast Two Hybrid technique does not exactly represent in vivo mammalian cell processes and these prey proteins may only interact with NFIB in an in vitro environment, rather than coming into proximity in vivo. Therefore, these interactions should be confirmed in mammalian cells.

Coimmunoprecipitation may be a good start for further confirming the interactions of identified proteins. For this, we can incubate mammalian cell or nuclear extracts with antibody coupled resin. After precipitating extracts and removing unbound proteins by washing, the elute can be analyzed in terms of containing interacted proteins. The interaction can then be resolved by SDS-PAGE and visualized by western blotting.

GST-pulldown technique can be used to confirm coimmunoprecipitation results or can be directly applied to study instead of coimmunoprecipitation. In this technique we can tag one of our bait proteins with glutathione S-transferase (GST) and capture it on an immobilized affinity ligand specific for the tag. Subsequently, we can incubate this immobilized bait with cell extracts. After washing away unbound proteins, we can elute interaction complex and analyze it by SDS-PAGE and western blotting.

A further step to prove interactions may be colocalization experiment. By applying fluorescence microscopy and via labelling two interacting proteins with fluorescent labels, we can observe the nearly localization of two proteins in the cell environment.

Finally, the significance of these interactions can be investigated in vivo by setting up mouse models.

REFERENCES

Adams, A.D., Choate, D.M., Thompson, M.A. (1995) NF1-L is the DNA-binding component of the protein complex at the peripherin negative regulatory element. J. Biol. Chem., 270, 6975–6983.

Alevizopoulos A., Dusserre Y., Pflugfelder MT., von der Weid T., Wahli W, Mermod N. (1995) A proline-rich TGF-β responsive transcriptional activator interacts with histone H3. GENES & DEVELOPMENT, 9, 3051-3066.

Altmann, H., Wendler, W., and Winnacker, E. L. (1994) Proc. Natl. Acad. Sci. U.S.A., 91, 3901–3905.

Asprer JS, Lee B, Wu CS, Vadakkan T, Dickinson ME, Lu HC, Lee SK. (2011) LMO4 functions as a co-activator of neurogenin 2 in the developing cortex. Development, 138(13), 2823-32.

Atabay KD, Karabay A (2012) Pin1 inhibition activates cyclin D and produces neurodegenerative pathology. J Neurochem.,120(3), 430-9.

Auerbach D and Stagljar I (2005) Yeast Two-Hybrid Protein–Protein Interaction Networks. Proteomics and Protein–Protein Interactions: Biology, Chemistry, Bioinformatics, and Drug Design, edited by Waksman. Springer, New York.

Bachurski CJ., Yang GH., Currier TA., Gronostajski RM., Hong D. (2003) Nuclear Factor I/Thyroid Transcription Factor 1 Interactions Modulate Surfactant Protein C Transcription.MOLECULAR AND CELLULAR BIOLOGY., p: 9014–9024.

Bandyopadhyay S. and Gronostajskiln RM (1994) Identification of a Conserved Oxidation-sensitive Cysteine Residue in the NFI Family of DNA-binding Proteins. THE JOURNAOLF BIOLOGICAL CHEMISTRY, 269(47), 29949-29955.

Barnabe-Heider F, Wasylnka JA, Fernandes KJ et al. (2005) Evidence that embryonic neurons regulate the onset of cortical gliogenesis via cardiotrophin-1. Neuron, 48, 253–265.

Barry, G., Piper,M, Lindwall,C., Moldrich,R., Mason, S., Little, E., Sarkar,A.,Tole,S., Gronostajski,RM., and. Richards LJ. (2008) Specific Glial Populations Regulate Hippocampal Morphogenesis The Journal of Neuroscience, 28(47),12328 –12340.

Bernard F, Gelsi-Boyer V, Murati A, Giraudier S, Trouplin V, Adélaïde J et al. (2009) Alterations of NFIA in chronic malignant myeloid diseases. Leukemia, 23, 583–585.

Bettencourt C. and Lima M. (2011) Machado-Joseph Disease: from first descriptions to new perspectives. Orphanet J Rare Dis., 6(35), 1150-1172.

Bettencourt C., Raposo M., Kazachkova N., Santos C., Kay T. et al. (2012) Sequence Analysis of 5′ Regulatory Regions of the Machado–Joseph Disease Gene (ATXN3). Cerebellum, 11(4),1045-50.

Brückner A., Polge C., Lentze N., Auerbach D., Schlattner U. (2009) Yeast Two-Hybrid, a Powerful Tool for Systems Biology *Int. J. Mol. Sci.*

Buschdorf JP, Chew LL, Soh UJ, Liou YC, Low BC (2008) Nerve growth factor stimulates interaction of Cayman ataxia protein BNIP-H/Caytaxin with peptidyl-prolyl isomerase Pin1 in differentiating neurons. PLoS One, 3(7),e2686.

Campbell CE, Piper M, Plachez C, Yeh Y-T, Baizer JS, Osinski JM, Litwack ED, Richards LJ, Gronostajski RM. (2008) The transcription factor Nfix is essential for normal brain development. BMC Dev Biol., 8,52.

Cato, A. C., Skroch, P., Weinmann, J., Butkeraitis, P., and Ponta, H. (1988) EMBO J., 7, 1403–1410.

Cederquist GY, Azim E, Shnider SJ, Padmanabhan H, Macklis JD (2013) Lmo4 establishes rostral motor cortex projection neuron subtype diversity. J Neurosci., 33(15), 6321-32.

Chang CY, Pasolli HA, Giannopoulou EG, Guasch G, Gronostajski RM, Elemento O, Fuchs E. (2013) NFIB is a governor of epithelial-melanocyte stem cell behaviour in a shared niche. Nature, 495(7439), 98-102.

Chaudhry AZ, Lyons GE, Gronostajski RM. (1997) Expression patterns of the four nuclear factor I genes during mouse embryogenesis indicate a potential role in development. Dev Dyn., 208, 313–325.

Chaudhry AZ., Vitullo AD., and Gronostajski RM. (1998) Nuclear Factor I (NFI) Isoforms Differentially Activate Simple versus Complex NFI-responsive Promoters. THE JOURNAL OF BIOLOGICAL CHEMISTRY., 273(29), 18538–18546.

Choi J, Ababon MR, Matteson PG, Millonig JH. (2011) Cut-like homeobox 1 and nuclear factor I/B mediate ENGRAILED2 autism spectrum disorder-associated haplotype function. Hum Mol Genet.. 1-15, doi:10.1093/hmg/ddr594.

Chu Z, Sun Y, Kuan CY, Grabowski GA, Qi X (2005) Saposin C: neuronal effect and CNS delivery by liposomes. Ann N Y Acad Sci.,1053:237-46.

Cooke DW, Lane, MD (1999) The transcription factor nuclear factor I mediates repression of the GLUT4 promoter by insulin. J. Biol. Chem. 274, 12917–12924.

Corbin JG, Gaiano N, Juliano SL, Poluch S, Stancik E, Haydar TF. (2008) Regulation of neural progenitor cell development in the nervous system. J. Neurochem., 106, 2272–2287.

Criekinge WV, Beyaert R. (1999) Yeast Two-Hybrid: State of the Art. Biological Procedures Online. Vol. 2 No. 1.

Dakson A, Yokota O, Esiri M, Bigio EH, Horan M et al. (2011) Granular expression of prolyl-peptidyl isomerase PIN1 is a constant and specific feature of Alzheimer's disease pathology and is independent of tau, Aβ and TDP-43 pathology. Acta Neuropathol.,121(5),635-49.

das Neves, L., Duchala, C., Godinho, F., Haxhiu, M., Colmenares, C., Macklin, W., Campbell, C., Butz, K., Gronostajski, R. (1999) Disruption of the murine Nuclear Factor I-A gene (Nfia) results in perinatal lethality, hydrocephalus and agenesis of the corpus callosum. Proc. Natl. Acad. Sci., 96, 11946–11951.

Deneen B, Ho R, Lukaszewicz A, Hochstim CJ, Gronostajski RM, Anderson DJ. (2006). The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. Neuron, 52,953–968 .

Doss-Pepe EW, Stenroos ES, Johnson WG, Madura K. (2003) Ataxin-3 interactions with rad23 and valosin-containing protein and its associations with ubiquitin chains and the proteasome are consistent with a role in ubiquitin-mediated proteolysis. Mol Cell Biol., 23(18), 6469-83.

Driller K, Pagenstecher A, Uhl M, Omran H, Berlis A, Gru¨ nder A, Sippel AE. (2007) Nuclear factor I X deficiency causes brain malformation and severe skeletal defects. Mol Cell Biol., 27,3855–3867.

Esnault G, Majocchi S, Martinet D, Besuchet-Schmutz N, Beckmann JS, Mermod N: (2009) Transcription factor CTF1 acts as a chromatindomain boundary that shields human telomeric genes from silencing. Mol Cell Biol, 29(9): 2409-2418.

Ferrari S, Simmen KC., Dusserre Y., Müller K, Fourel G, Gilson E, Mermod N. (2004) Chromatin Domain Boundaries Delimited by a Histone-binding Protein in Yeast, 279(53), 55520–55530.

Ferro A, Carvalhoc AL, Teixeira-Castroe A, Almeidad C, Toméf RJ et al. (2007) NEDD8: A new ataxin-3 interactor. Biochim Biophys Acta., 1773(11), 1619-27.

Fields S, Song O. (1989) A novel genetic system to detect protein-protein interactions. Nature, 340, 245-246.

Fujita S, Ito T, Mizutani T, Minoguchi S, Yamamichi N, Sakurai K, Iba H. (2008) miR-21 Gene expression triggered by AP-1 is sustained through a double-negative feedback mechanism. J. Mol. Biol., 378, 492-504.

Fujiwara Y, Kasashima K, Saito K, Fukuda M, Fukao A et al. (2011) Microtubule association of a neuronal RNA-binding protein HuD through its binding to the light chain of MAP1B. Biochimie., 93(5), 817-22.

Gao, B., Jiang, L., Kunos, G. (1996) Transcriptional regulation of alpha(1b) adrenergic receptors (alpha(1b)AR) by nuclear factor 1 (NF1): a decline in the concentration of NF1 correlates with the

downregulation of alpha(1b)AR gene expression in regenerating liver. Mol. Cell. Biol., 16, 5997–6008.

Gesemann M, Litwack ED, Yee KT, Christen U, O'Leary DD (2001) Identification of candidate genes for controlling development of the basilar pons by differential display PCR. Mol Cell Neurosci., 18, 1–12.

Geurts, J.M., Schoenmakers, E.F., Roijer, E., Astrom, A.K., Stenman, G., van de Ven, W.J. (1998) Identification of NFIB as recurrent translocation partner gene of HMGIC in pleomorphic adenomas. Oncogene, 16, 865–872.

Gomi F, Uchida Y. (2012) MAP1B 1-126 interacts with tubulin isoforms and induces neurite outgrowth and neuronal death of cultured cortical neurons. Brain Res., 1433,1-8.

Gounari, F., Francesco, RD, Schmitt, J.,.van der VIiet PC., Cortese R., and Stunnenberg H. (1990) Amino-terminal domain of NF1 binds to DNA as a dimer and activates adenovirus DNA replication. The EMBO Journal,.9(2), 559-566.

Graves, R.A., Tontonoz, P., Ross, S.R., Spiegelman, B.M. (1991) Identification of a potent adipocyte-specific enhancer: involvement of an NF-1-like factor. Genes Dev., 5, 428–437.

Grewal, T., Theisen, M., Borgmeyer, U., Grussenmeyer, T., Rupp, R. A. W., et al. (1992) *Mol. Cell. Biol.*, 12, 2339–2350.

Gronostajski, R.M., Adhya, S., Nagata, K., Guggenheimer, R.A., Hurwitz, J., (1985). Site-specific DNA binding of nuclear factor I: Analyses of cellular binding sites. Mol. Cell. Biol., 5, 964–971.

Gronostajski RM. (2000) Roles of the NFI/CTF gene family in transcription and development. Gene, 249, 31–45.

Gronostajski RM, Guaneri J, Lee DH and Gallo SM. (2011) The NFI-Regulome Database: A tool for annotation and analysis of control regions of genes regulated by Nuclear Factor I transcription factors. Journal of Clinical Bioinformatics, 1:4

Guillemot, F (2007a) Cell fate specification in the mammalian telencephalon. Prog. Neurobiol. 83, 37-52.

Guillemot F (2007b) Spatial and temporal specification of neural fates by transcription factor codes Development, 134, 3771-3780 doi:10.1242/dev.006379

Guo F, Huang X, Li S, Sun L, Li Y, Li H, Zhou Y, Chu Y, Zhou T (2007) Identification of prosaposin as a novel interaction partner for Rhox5. J Genet Genomics, 34(5),392-9.

Haacke, A., Broadley, S. A., Boteva, R., Tzvetkov, N., Hartl, F. U., Breuer, P. (2006) Proteolytic cleavage of polyglutamine-expanded ataxin-3 is critical for aggregation and sequestration of non-expanded ataxin-3. Hum. Molec. Genet., 15, 555-568.

Hahm K, Sum EY, Fujiwara Y, Lindeman GJ, Visvader JE, Orkin SH (2004) Defective neural tube closure and anteroposterior patterning in mice lacking the LIM protein LMO4 or its interacting partner Deaf-1. Mol Cell Biol., 24(5), 2074-82.

Hennighausen, L., Siebenlist, U., Danner, D., Leder, P., Rawlins, D., Rosenfeld, P., Kelly, T.J., (1985). High-affinity binding site for a specific nuclear protein in the human IgM gene. Nature, 314, 289–292.

Jackson, D. A., Rowader, K. E., Stevens, K., Jiang, C., Milos, P., and Zaret, K. S. (1993) Mol. Cell. Biol., 13, 2401–2410.

Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J. & Tjian, R. (1987). A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell, 48, 79-89.

Joshi K, Lee S, Lee B, Lee JW, Lee SK (2009) LMO4 controls the balance between excitatory and inhibitory spinal V2 interneurons. Neuron, 61(6), 839-51

Kang P, Lee HK, Glasgow SM, Finley M, Donti T et al. (2012) Sox9 and NFIA coordinate a transcriptional regulatory cascade during the initiation of gliogenesis. Neuron, 74(1), 79-94.

Kashani AH, Qiu Z, Jurata L, Lee SK, Pfaff S, Goebbels S, Nave KA, Ghosh A (2006) Calcium activation of the LMO4 transcription complex and its role in the patterning of thalamocortical connections. J Neurosci., 26(32), 8398-408.

Kenny DA,, Jurata LW, Saga Y, and Gill GN (1998) Identification and characterization of LMO4, an LMO gene with a novel pattern of expression during embryogenesis. PNAS. 95(19), 11257-11262.

Koehler U, Holinski-Feder E, Ertl-Wagner B, Kunz J, von Moers A, von Voss

H, Schell-Apacik C. (2010) A novel 1p31.3p32.2 deletion involving the NFIA gene detected by arrayCGHin a patient with macrocephaly and hypoplasia of the corpus callosum. Eur J Pediatr., 169,463–468.

Kruse, U., Qian, F., and Sippel, AE (1991) Identification of a fourth nuclear factor I gene in chicken by cDNA cloning: NFI-X. Nucleic Acids Res., 19, 6641.

Kruse, U., and Sippel, AE (1994a) The genes for transcription factor nuclear factor I give rise to corresponding splice variants between vertebrate species. J. Mol. Biol., 238, 860–865.

Kruse, U., and Sippel, AE (1994b) Transcription factor nuclear factor I proteins form stable homo- and heterodimers. FEBS Lett., 348,46.

Leahy P, Crawford DR, Grossman G,. Gronostajski RM and Hanson RM (1999) CREB Binding Protein Coordinates the Function of Multiple Transcription Factors Including Nuclear Factor I to Regulate Phosphoenolpyruvate Carboxykinase (GTP) Gene Transcription. THE JOURNAL OF BIOLOGICAL CHEMISTRY, 274(13), 8813–8822.

Lee SK, Jurata LW, Nowak R, Lettieri K, Kenny DA, Pfaff SL, Gill GN. (2005) The LIM domain-only protein LMO4 is required for neural tube closure. Mol Cell Neurosci., 28(2), 205-14.

Leegwater, P.A.J., van Driel, W., van der Vliet, P.C., (1985) Recogni tion site of nuclear factor I, a sequence-specific DNA-binding pro tein from HeLa cells that stimulates adenovirus DNA replication. EMBO J., 4, 1515–1521.

Leuba G, Vernay A, Vu D, Walzer C, Belloir B et al. (2004) Differential expression of LMO4 protein in Alzheimer's disease. Neuropathol Appl Neurobiol., 30(1), 57-69.

Li N, Sarojini H, An J, Wang E (2010) Prosaposin in the secretome of marrow stroma-derived neural progenitor cells protects neural cells from apoptotic death. J Neurochem.112(6),1527-38.

Lu W, Quintero-Rivera F, Fan Y, Alkuraya FS, Donovan DJ, Xi Q, et al. (2007) NFIA haploinsufficiency is associated with a CNS malformation syndrome and urinary tract defects. PLoS Genet.,3:e80.

Macleod, K., Plumb, M. (1991) Derepression of mouse β-major-globin gene transcription during erythroid differentiation. Mol. Cell. Biol., 11, 4324–4332.

Manetopoulos C, Hansson A, Karlsson J, Jönsson JI, Axelson H (2003) The

LIM-only protein LMO4 modulates the transcriptional activity of HEN1. Biochem Biophys Res Commun., 307(4), 891-9.

Mason S, Piper M, Gronostajski RM, Richards LJ. (2009) Nuclear factor one transcription factors in CNS development. Mol Neurobiol., 39:10e23.

Meisterernst, M., Gander, I., Rogge, L., Winnacker, E.L. (1988). A quantitative analysis of nuclear factor I/DNA interactions. Nucleic Acids Res., 16, 4419–4435.

Mermod, N.O, Neill, E., Kelly, T., Tjian, R., (1989). The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. Cell, 58, 741–753.

Miller FD, Gauthier AS (2007) Timing is everything: Making neurons versus glia in the developing cortex. Neuron, 54,357–369.

Mitani Y, Li J, Rao PH, Zhao Y et al. (2010) Comprehensive Analysis of the MYB-NFIB Gene Fusion in Salivary Adenoid Cystic Carcinoma: Incidence, Variability, and Clinicopathologic Significance. Clin Cancer Res, 16,4722-4731.

Mizutani K, Saito T (2005) Progenitors resume generating neurons after temporary inhibition of neurogenesis by Notch activation in the mammalian cerebral cortex. Development, 132, 1295–1304.

Monaci, P., Nuzzo, M., Stampfli, S., Tollervey, D., De Simone, V., Nicosia, A. (1995) A complex interplay of positive and negative elements is responsible for the different transcriptional activity of liver NF1 variants. Mol. Biol. Rep., 21, 147–158.

Moon H, Hwang K, Kim J, Kim HS, Lee M et al. (2011) NFIB is a potential target for estrogen receptor-negative breast cancers. MOLECULAR ONCOLOGY, 5,538-544.

Moritz A, Scheschonka A, Beckhaus T, Karas M, Betz H (2009) Metabotropic glutamate receptor 4 interacts with microtubule-associated protein 1B. Biochem Biophys Res Commun., 390(1),82-6.

Nagata, K., Guggenheimer, R., Enomoto, T., Lichy, J., Hurwitz, J. (1982). Adenovirus DNA replication in vitro identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. Proc. Natl. Acad. Sci.,79, 6438–6442.

Nakamura K, Kosugi I, Lee DY, Hafner A, Sinclair DA, Ryo A, Lu KP (2012) Prolyl isomerase Pin1 regulates neuronal differentiation via β-catenin. Mol Cell Biol.,32(15), 2966-78.

Namihira M, Kohyama J, Semi K, Sanosaka T, Deneen B, Taga T, Nakashima K. (2009) Committed neuronal precursors confer astrocytic potential on residual neural precursor cells. Dev Cell., 16, 245–255.

Nebl, G., Cato, A. (1995) NFI/X proteins: a class of NFI family of transcription factors with positive and negative regulatory domains. Cell. Mol. Biol. Res., 41, 85–95.

Nowock, J., Borgmeyer, U., Puschel, A.W., Rupp, R.A.W., Sippel, A.E. (1985). The TGGCA protein binds to the MMTV-LTR, the adenovirus origin of replication, and the BK virus enhancer. Nucleic Acids Res., 13, 2045–2061.

Ochiai W, Yanagisawa M, Takizawa T et al. (2001) Astrocyte differentiation of fetal neuroepithelial cells involving cardiotrophin-1-induced activation of STAT3. Cytokine, 14, 264–271.

Osada, S., Daimon, S., Ikeda, T., Nishihara, T., Yano, K., Yamasaki, M., Imagawa, M. (1997a) Nuclear factor 1 family proteins bind to the silencer element in the rat glutathione transferase P gene. J. Biochem., 121, 355–363.

Osada, S., Ikeda, T., Xu, M., Nishihara, T., Imagawa, M. (1997b) Identification of the transcriptional repression domain of nuclear factor 1-A. Biochem. Biophys. Res. Commun., 238, 744–747.

Paonessa, G., Gounari, F., Frank, R., and Cortese, R. (1988) EMBO J., 7, 3115–3123.

Park JC, Herr Y, Kim HJ, Gronostajski RM, Cho MI. (2007) Nfic gene disruption inhibits differentiation of odontoblasts responsible for root formation and results in formation of short and abnormal roots in mice. J Periodontol., 78, 1795–1802.

Paulson HL, Das SS, Crino PB, Perez MK, Patel SC et al. (1997) Machado-Joseph disease gene product is a cytoplasmic protein widely expressed in brain. Ann Neurol., 41(4),453-62.

Persson M, Andren Y, Mark J, Horlings HM, Persson F, Stenman G. (2009) Recurrent fusion of MYB and NFIB transcription factor genes in

carcinomas of the breast and head and neck. Proc Natl Acad Sci U S A,106. 18740–4.

Piper M, Moldrich RX, Lindwall C, Little E, Barry G et al. (2009) Multiple non-cell-autonomous defects underlie neocortical callosal dysgenesis in Nfib-deficient mice. Neural Dev., 4, 43.

Piper M, Barry G, Hawkins J et al. (2010) NFIA Controls Telencephalic Progenitor Cell Differentiation through Repression of the Notch Effector Hes1. The Journal of Neuroscience, 30(27), 9127–9139.

Plachez C, Lindwall C, Sunn N, Piper M, Moldrich RX, Campbell CE, et al. (2008) Nuclear factor I gene expression in the developing forebrain. J Comp Neurol., 508, 385e401.

Plasari G, Edelmann S, Högger F, Dusserre Y, Mermod N, and Calabrese A. (2010) Nuclear Factor I-C Regulates TGF-ß-dependent Hair Follicle Cycling. The Journal Of Biologıcal Chemistry.,285(44), 34115–34125.

Rajas, F., Delhase, M., De La Hoya, M., Verdood, P., Castrillo, J.L., Hooghe-Peters, E.L. (1998) Nuclear factor 1 regulates the distal silencer of the human PIT1/GHF1 gene. Biochem. J., 333, 77–84.

Rowitch, D. H (2004) Glial specification in the vertebrate neural tube. Nat. Rev.Neurosci., 5, 409-419.

Ryo A, Togo T, Nakai T, Hirai A, Nishi M et al. (2006) Prolyl-isomerase Pin1 accumulates in lewy bodies of parkinson disease and facilitates formation of alpha-synuclein inclusions. J Biol Chem., 281(7), 4117-25.

Sadreyev R, Grishin N. (2003) COMPASS: A Tool for Comparison of Multiple Protein Alignments with Assessment of Statistical Significance. Journal of Molecular Biology, 326(1), 317–336.

Santoro, C., Mermod, N., Andrews, P. C., and Tjian, R. (1988) Nature, 334, 218–224.

Schaffar G, Taniguchi J, Brodbeck T, Meyer AH, Schmidt M, Yamashita T, Mueller BK (2008) LIM-only protein 4 interacts directly with the repulsive guidance molecule A receptor Neogenin. J Neurochem., 107(2),418-31.

Schock SC, Xu J, Duquette PM, Qin Z, Lewandowski AJ et al. (2008) Rescue of neurons from ischemic injury by peroxisome proliferator-activated receptor-gamma requires a novel essential cofactor LMO4. J Neurosci., 28(47), 12433-44.

Schuur, E.R., Kruse, U., Iacovoni, J.S., Vogt, P.K. (1995) Nuclear factor I interferes with transformation induced by nuclear oncogenes. Cell Growth Differ., 6, 219–227.

Selcuklu, S.D., Donoghue, M.T., Spillane, C. (2009) miR-21 as a key regulator of oncogenic processes. Biochem. Soc. Trans., 37, 918-925.

Shu T, Butz KG, Plachez C, Gronostajski RM, Richards LJ. (2003). Abnormal development of forebrain midline glia and commissural projections in Nfia knock-out mice. J Neurosci., 23, 203–212.

Sikora J, Harzer K, Elleder M. (2007) Neurolysosomal pathology in human prosaposin deficiency suggests essential neurotrophic function of prosaposin. Acta Neuropathol., 113(2),163-75.

Singh, M.V., Ntambi, J.M., (1998) Nuclear factor 1 is essential for the expression of stearoyl-CoA desaturase 1 gene during preadipocyte differentiation. Biochim. Biophys. Acta., 1398, 148–156.

Singh U, Bongcam-Rudloff E, Westermark B (2009) A DNA Sequence Directed Mutual Transcription Regulation of HSF1 and NFIX Involves Novel Heat Sensitive Protein Interactions. PLoS ONE, 4(4), e5050.

Steele-Perkins G, Butz KG, Lyons GE, et al. (2003) Essential role for NFI-C/CTF transcription-replication factor in tooth root development. Mol Cell Biol, 23, 1075–1084.

Steele-Perkins G, Plachez C, Butz KG, Yang G, Bachurski CJ, Kinsman SL, Litwack ED, Richards LJ, Gronostajski RM. (2005) The transcription factor gene Nfib is essential for both lung maturation and brain development.Mol Cell Biol., 25, 685– 698.

Stevens, B., Allen, N. J., Vazquez, L. E., Howell, G. R., Christopherson, K. S. et al. (2007) The classical complement cascade mediates CNS synapse elimination. Cell., 131,1164-1178.

Szabo, P., Moitra, J., Rencendorj, A., Rakhely, G., Rauch, T., Kiss, I. (1995) Identification of a nuclear factor-I family protein-binding site in the silencer region of the cartilage matrix protein gene. J. Biol. Chem., 270, 10212–10221.

Tanese N, Pugh BF, and Tjian R (1991) Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. GENES & DEVELOPMENT, 5, 2212-2224.

Trevaskis, J., Walder, K., Foletta, V., Kerr-Bayles, L., McMillan, J., et al. (2005) Src homology 3-domain growth factor receptor-bound 2-like (endophilin) interacting protein 1, a novel neuronal protein that regulates energy balance. Endocrinology, 146, 3757-3764.

Tsai-Pflugfelder M, Gasser SM, and Wahli W (1998) Functional Interaction between the Estrogen Receptor and CTF1: Analysis of the Vitellogenin Gene B1 Promoter in Yeast. MOL END., 12(10), 1525-1541.

Tse E, Smith AJ, Hunt S, Lavenir I, Forster A et al. (2004) Null mutation of the Lmo4 gene or a combined null mutation of the Lmo1/Lmo3 genes causes perinatal lethality, and Lmo4 controls neural tube development in mice. Mol Cell Biol.,24(5),2063-73.

Uezu, A., Horiuchi, A., Kanda, K., Kikuchi, N., Umeda, K., et al. (2007) SGIP1-alpha is an endocytic protein that directly interacts with phospholipids and Eps15. J. Biol. Chem., 282, 26481-26489.

Utreras E, Jiménez-Mateos EM, Contreras-Vallejos E, Tortosa E, Pérez M et al. (2008) Microtubule-associated protein 1B interaction with tubulintyrosine ligase contributes to the control of microtubule tyrosination. Dev Neurosci., 30(1-3), 200-10.

Xiao H, Lis JT, Xiao H, Greenblatt J and Friesen JD (1994) The upstream activator CTF/NF1 and RNA polymerase II share a common element involved in transcriptional activation. Nucleic Acids Research, 22(11), 1966-1973.

Waki H, Nakamura M, Yamauchi T, Wakabayashi K-i, Yu J, et al. (2011) Global Mapping of Cell Type–Specific Open Chromatin by FAIRE-seq Reveals the Regulatory Role of the NFI Family in Adipocyte Differentiation. PLoS Genet., 7(10), e1002311, doi:10.1371/journal.pgen.1002311

Wang G, Sawai N, Kotliarova S, Kanazawa I, Nukina N. (2000) Ataxin-3, the MJD1 gene product, interacts with the two human homologs of yeast DNA repair protein RAD23, HHR23A and HHR23B. Hum Mol Genet., 9(12),1795-803.

Wang W, Mullikin-Kilpatrick D, Crandall JE, Gronostajski RM, Litwack ED, Kilpatrick DL. (2007) Nuclear factor I coordinates multiple phases of cerebellar granule cell development via regulation of cell adhesion molecules. J Neurosci., 27,6115– 6127.

Wilczynska, K. M., Singh, S. K., Adams, B., Bryan, L., Rao, R. R. et al. (2009) Stem Cells, 27, 1173–1181.

Wuarin, J., Mueller, C. & Schibler, U. (1990). A ubiquitous CCAAT factor is required for efficient in vitro transcription from the mouse albumin promoter. J. Mol. Biol., 214, 865-874.

Xu C, Deng F, Mao Z, Zhang J, Wang H et al. (2012) The interaction of the second Kunitz-type domain (KD2) of TFPI-2 with a novel interaction partner, prosaposin, mediates the inhibition of the invasion and migration of human fibrosarcoma cells. Biochem J., 441(2), 665-74.

Yoneshige A, Suzuki K, Suzuki K, Matsuda J (2010) A mutation in the saposin C domain of the sphingolipid activator protein (Prosaposin) gene causesneurodegenerative disease in mice. J Neurosci Res.,88(10), 2118-34.

Url-1 *<http://www.genecards.org/cgi-bin/carddisp.pl?gene=LMO4>*

Url-2 *<http://www.genecards.org/cgi-bin/carddisp.pl?gene=PIN1>*

APPENDICES

APPENDIX A: Sequence Data for NFIB2 and NFIB2Δ Inserts
APPENDIX B: Sequence Alignments of Prey Inserts

APPENDIX A

```
   1 ATGATGTATT CTCCCATCTG TCTCACTCAG GATGAATTCC ACCCATTTAT TGAGGCACTT
  61 CTTCCTCACG TCCGTGCAAT TGCCTATACT TGGTTCAACC TGCAGGCTCG AAAACGCAAG
 121 TACTTTAAAA AGCATGAGAA ACGAATGTCG AAGGATGAAG AAAGGGCAGT CAAAGACGAG
 181 CTGCTCAGTG AGAAGCCCGA AATCAAGCAG AAGTGGGCAT CCAGGCTCCT GGCCAAACTG
 241 CGCAAAGATA TCCGCCAGGA GTACCGGGAG GACTTTGTGC TTACCGTGAC TGGCAAGAAG
 301 CACCCGTGCT GTGTCTTATC CAATCCAGAC CAGAAGGGTA AGATTAGGAG GATCGACTGC
 361 CTGCGACAGG CAGACAAAGT CTGGCGTCTG GATCTAGTCA TGGTGATCCT GTTCAAAGGC
 421 ATCCCTTTGG AGAGTACGGA TGGAGAGCGA CTCATGAAGT CCCCGCACTG CACAAACCCA
 481 GCACTTTGTG TTCAGCCACA CCACATCACA GTATCAGTTA AGGAGCTTGA CTTGTTTTTG
 541 GCATACTACG TGCAGGAGCA AGATTCTGGA CAATCAGGAA GTCCAAGCCA CAGTGATCCT
 601 GCCAAGAATC CTCCAGGGTA CCTCGAGGAC AGCTTTGTAA AATCCGGAGT CTTCAATGTA
 661 TCAGAGCTTG TGAGAGTATC CAGAACACCC ATAACCCAGG GAACTGGAGT CAACTTCCCA
 721 ATCGGAGAAA TTCCCAGCCA ACCATACTAT CATGACATGA ACTCTGGTGT GAACCTGCAG
 781 AGGTCGCTGT CTTCTCCACC GAGCAGCAAA AGACCCAAAA CTATATCTAT AGATGAAAAT
 841 ATGGAGCCAA GTCCTACAGG AGACTTTTAC CCCTCTCCAA ATTCACCAGC TGCTGGAAGT
 901 CGAACATGGC ATGAACGAGA TCAAGATATG TCTTCTCCAA CTACAATGAA GAAGCCTGAG
 961 AAGCCACTGT TTAGCTCTAC ATCTCCACAG GATTCTTCCC CAAGATTGAG CACTTTCCCC
1021 CAGCACCATC ATCCCGGAAT ACCTGGAGTC GCGCACAGTG TCATCTCAAC TCGAACTCCA
1081 CCTCCGCCCT CACCGTTGCC ATTTCCGACG CAAGCTATCC TTCCTCCGGC ACCTTCCAGC
1141 TACTTCTCTC ATCCAACAAT CAGATATCCT CCTCACCTGA ATCCTCAGGA TACTCTGAAG
1201 AACTACGTAC CTTCTTATGA CCCATCCAGT CCTCAAACGA GCCAGTCCTG GTACCTGGGC
1261 TAG
```

Figure A.1 Sequence data for NFIB2 insert.

```
   1 GAGGACAGCT TTGTAAAATC CGGAGTCTTC AATGTATCAG AGCTTGTGAG AGTATCCAGA
  61 ACACCCATAA CCCAGGGAAC TGGAGTCAAC TTCCCAATCG GAGAAATTCC CAGCCAACCA
 121 TACTATCATG ACATGAACTC TGGTGTGAAC CTGCAGAGGT CGCTGTCTTC TCCACCGAGC
 181 AGCAAAAGAC CCAAAACTAT ATCTATAGAT GAAAATATGG AGCCAAGTCC TACAGGAGAC
 241 TTTTACCCCT CTCCAAATTC ACCAGCTGCT GGAAGTCGAA CATGGCATGA ACGAGATCAA
 301 GATATGTCTT CTCCAACTAC AATGAAGAAG CCTGAGAAGC CACTGTTTAG CTCTACATCT
 361 CCACAGGATT CTTCCCCAAG ATTGAGCACT TTCCCCCAGC ACCATCATCC CGGAATACCT
 421 GGAGTCGCGC ACAGTGTCAT CTCAACTCGA ACTCCACCTC CGCCCTCACC GTTGCCATTT
 481 CCGACGCAAG CTATCCTTCC TCCGGCACCT TCCAGCTACT TCTCTCATCC AACAATCAGA
 541 TATCCTCCTC ACCTGAATCC TCAGGATACT CTGAAGAACT ACGTACCTTC TTATGACCCA
 601 TCCAGTCCTC AAACGAGCCA GTCCTGGTAC CTGGGCTAG
```

Figure A.2 Sequence data for NFIB2Δ insert.

APPENDIX B

Homo sapiens LIM domain only 4 (LMO4), mRNA

```
Score = 1685 bits(912), Expect = 0.0, Identities = 21/925(99%)
Gaps = 2/925(0%),Strand = Plus/Plus

Query  102   GCGATTT-aaaaaaaaaaaaaaGCCGcccttagccccctcctcccctttcctgcttctgc  160
             ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   596   GCGATTTAAAAAAAAAAAAAAAAGCCGCCCTTAGCCCCCTCCTCCCCTTTCCTGCTTCTGC  655

Query  161   gagaactccctccctccctccagctccGCCAGCCCAGGCGCCCCTTCCCTGGAAGCCGAG  220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   656   GAGAACTCCCTCCCTCCCTCCAGCTCCGCCAGCCCAGGCGCCCCTTCCCTGGAAGCCGAG  715

Query  221   CGGCTTCGCTCGCATTTCACCGCCGCCGCCTCTCGCAATATTGCAATATAGGGGAAAAGC  280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   716   CGGCTTCGCTCGCATTTCACCGCCGCCGCCTCTCGCAATATTGCAATATAGGGGAAAAGC  775

Query  281   AGACCATGGTGAATCCGGGCAGCAGCTCGCAGCCGCCCCCGGTGACGGCCGGCTCCCTCT  340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   776   AGACCATGGTGAATCCGGGCAGCAGCTCGCAGCCGCCCCCGGTGACGGCCGGCTCCCTCT  835

Query  341   CCTGGAAGCGGTGCGCAGGCTGCGGGGGCAAGATTGCGGACCGCTTTCTGCTCTATGCCA  400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   836   CCTGGAAGCGGTGCGCAGGCTGCGGGGGCAAGATTGCGGACCGCTTTCTGCTCTATGCCA  895

Query  401   TGGACAGCTATTGGCACAGCCGGTGCCTCAAGTGCTCCTGCTGCCAGGCGCAGCTGGGCG  460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   896   TGGACAGCTATTGGCACAGCCGGTGCCTCAAGTGCTCCTGCTGCCAGGCGCAGCTGGGCG  955

Query  461   ACATCGGCACGTCCTGTTACACCAAAAGTGGCATGATCCTTTGCAGAAATGACTACATTA  520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   956   ACATCGGCACGTCCTGTTACACCAAAAGTGGCATGATCCTTTGCAGAAATGACTACATTA  1015

Query  521   GGTTATTTGGAAATAGCGGTGCTTGCAGCGCTTGCGGACAGTCGATTCCTGCGAGTGAAC  580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   1016  GGTTATTTGGAAATAGCGGTGCTTGCAGCGCTTGCGGACAGTCGATTCCTGCGAGTGAAC  1075

Query  581   TCGTCATGAGGGCGCAAGGCAATGTGTATCATCTTAAGTGTTTTACATGCTCTACCTGCC  640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   1076  TCGTCATGAGGGCGCAAGGCAATGTGTATCATCTTAAGTGTTTTACATGCTCTACCTGCC  1135

Query  641   GGAATCGCCTGGTCCCGGGAGATCGGTTTCACTACATCAATGGCAGTTTATTTTGTGAAC  700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   1136  GGAATCGCCTGGTCCCGGGAGATCGGTTTCACTACATCAATGGCAGTTTATTTTGTGAAC  1195

Query  701   ATGATAGACCTACAGCTCTCATCAATGGCCATTTGAATTCACTTCAGAGCAATCCACTAC  760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   1196  ATGATAGACCTACAGCTCTCATCAATGGCCATTTGAATTCACTTCAGAGCAATCCACTAC  1255

Query  761   TGCCAGACCAGAAGGTCCGCTAAAAGGTCAGAGTAATGCAGAATGCGTGCCTTCATCTCA  820
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
LMO4   1256  TGCCAGACCAGAAGGTCTGCTAAAAGGTCAGAGTAATGCAGAATGCGTGCCTTCATCTCA  1315

Query  821   GATTTGTTCATCACAGGTGGATCCCATGTGTCTTCAGTAGACAAGTCACCTTTGTAGCTA  880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   1316  GATTTGTTCATCACAGGTGGATCCCATGTGTCTTCAGTAGACAAGTCACCTTTGTAGCTA  1375

Query  881   GCACCAGTGCCAGCTCCATGCCATTGCACCTTCTTTAGTCTTGATTGCCCTTCCCGCATT  940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   1376  GCACCAGTGCCAGCTCCATGCCATTGCACCTTCTTTAGTCTTGATTGCCCTTCCCGCATT  1435

Query  941   TATTGGTGTATTAAAATGACTGAATATGAACATTAAGGACTCCATGAACCTGGGCTAATG  1000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LMO4   1436  TATTGGTGTATTAAAATGACTGAATATGAACATTAAGGACTCCATGAACCTGGGCTAATG  1495

Query  1001  GGAGACTGTAGAGaaaaa-aaaaaa  1024
             ||||||||||||||||  ||||||
LMO4   1496  GGAGACTGTAGAGAAAATGAAAAAA  1520
```

Figure B.1 Alignment data for LMO4 mRNA.

This insert sequence is homologous to nt 596-1520 of LIM domain only 4 (LMO4), mRNA. (NM_006769.3).. This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for the first 164 amino acids of 165 amino acid LMO4 protein.

Homo sapiens prosaposin (PSAP), transcript variant 3, mRNA

```
Score = 1753 bits(949), Expect = 0.0,Identities = 969/977(99%)
Gaps = 8/977(0%),Strand = Plus/Plus

Query  105   AACCTGGAGAAAAACAGCACCAAGCAGGAGATCCTGGCTGCTCTTGAGAAAGGCTGCAGC  164
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1374  AACCTGGAGAAAAACAGCACCAAGCAGGAGATCCTGGCTGCTCTTGAGAAAGGCTGCAGC  1433

Query  165   TTCCTGCCAGACCCTTACCAGAAGCAGTGTGATCAGTTTGTGGCAGAGTACGAGCCCGTG  224
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1434  TTCCTGCCAGACCCTTACCAGAAGCAGTGTGATCAGTTTGTGGCAGAGTACGAGCCCGTG  1493

Query  225   CTGATCGAGATCCTGGTGGAGGTGATGGATCCTTCCTTCGTGTGCTTGAAAATTGGAGCC  284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1494  CTGATCGAGATCCTGGTGGAGGTGATGGATCCTTCCTTCGTGTGCTTGAAAATTGGAGCC  1553

Query  285   TGCCCCTCGGCCCATAAGCCCTTGTTGGGAACTGAGAAGTGTATATGGGGCCCAAGCTAC  344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1554  TGCCCCTCGGCCCATAAGCCCTTGTTGGGAACTGAGAAGTGTATATGGGGCCCAAGCTAC  1613

Query  345   TGGTGCCAGAACACAGAGACAGCAGCCCAGTGCAATGCTGTCGAGCATTGCAAACGCCAT  404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1614  TGGTGCCAGAACACAGAGACAGCAGCCCAGTGCAATGCTGTCGAGCATTGCAAACGCCAT  1673

Query  405   GTGTGGAACTAGGAGGAGGAATATTCCATCTTGGCAGAAACCACAGCATTGGtttttttC  464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1674  GTGTGGAACTAGGAGGAGGAATATTCCATCTTGGCAGAAACCACAGCATTGGTTTTTTTC  1733

Query  465   TACTTGTGTGTCTGGGGGAATGAACGCACAGATCTGTTTGACTTTGTTATAAAAATAGGG  524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1734  TACTTGTGTGTCTGGGGGAATGAACGCACAGATCTGTTTGACTTTGTTATAAAAATAGGG  1793

Query  525   CTCCCCCACCTCCCCCATTTCTGTGTCCTTTATTGTAGCATTGCTGTCTGCAAGGGAGCC  584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1794  CTCCCCCACCTCCCCCATTTCTGTGTCCTTTATTGTAGCATTGCTGTCTGCAAGGGAGCC  1853

Query  585   CCTAGCCCCTGGCAGACATAGCTGCTTCAGTGCCCCTTTTCTCTCTGCTAGATGGATGTT  644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1854  CCTAGCCCCTGGCAGACATAGCTGCTTCAGTGCCCCTTTTCTCTCTGCTAGATGGATGTT  1913

Query  645   GATGCACTGGAGGTCTTTTAGCCTGCCCTTGCATGGCGCCTGCTGGAGGAGGAGAGAGCT  704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1914  GATGCACTGGAGGTCTTTTAGCCTGCCCTTGCATGGCGCCTGCTGGAGGAGGAGAGAGCT  1973

Query  705   CTGCTGGCATGAGCCACAGTTTCTTGACTGGAGGCCATCAACCCTCTTGGTTGAGGCCTT  764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   1974  CTGCTGGCATGAGCCACAGTTTCTTGACTGGAGGCCATCAACCCTCTTGGTTGAGGCCTT  2033

Query  765   GTTCTGAGCCCTGACATGTGCTTGGGCACTGGTGGGCCTGGGCTTCTGAGGTGGCCTCCT  824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   2034  GTTCTGAGCCCTGACATGTGCTTGGGCACTGGTGGGCCTGGGCTTCTGAGGTGGCCTCCT  2093

Query  825   GCCCTGATCAGGGACCCTCCCCGCTTTCCTGGGCCTCTCAGTTGAACAAAGCAGCAAAAC  884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   2094  GCCCTGATCAGGGACCCTCCCCGCTTTCCTGGGCCTCTCAGTTGAACAAAGCAGCAAAAC  2153

Query  885   AAAGGCAGTTTTATATGAAAGATTAGAAGCCTGGAATAATCAGGCTTTTTAAATGATGTA  944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   2154  AAAGGCAGTTTTATATGAAAGATTAGAAGCCTGGAATAATCAGGCTTTTTAAATGATGTA  2213

Query  945   ATTCCCACTGTAATAGCATAGGGATTTTGGAAGCAGCTGCTGGTGGCTTGGGACATCAGT  1004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PSAP   2214  ATTCCCACTGTAATAGCATAGGGATTTTGGAAGCAGCTGCTGGTGGCTTGGGACATCAGT  2273

Query  1005  GGGGCCAAGG-TTCTCTGTCCCTGGT-CAACTGTGATT-GGCTTTCCCGTG-CTTTC-TG  1059
             |||||||||| |||||||||||||| |||||||||||| |||||||||||| ||||| ||
```

```
PSAP   2274  GGGGCCAAGGGTTCTCTGTCCCTGGTTCAACTGTGATTTGGCTTTCCCGTGTCTTTCCTG  2333

Query  1060  GTGA-GC-TTGTT-GGG  1073
             |||| || ||||| |||
PSAP   2334  GTGATGCCTTGTTTGGG  2350
```

Figure B.2 Alignment data for PSAP mRNA.

This insert sequence is homologous to nt 1374-2350 of Proactivator polypeptide isoform b preproprotein (PSAP) mRNA (NM_001042465.1). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for the last 103 amino acids of 526 amino acid PSAP protein.

Homo sapiens ataxin 3 (ATXN3), transcript variant o, mRNA

```
Score = 652 bits(353),Expect = 0.0,Identities = 353/353(100%),
Gaps = 0/353(0%),Strand = Plus/Plus

Query  107  ggCGTGGGGGCCGTTGGCTCCAGACAAATAAACATGGAGTCCATCTTCCACGAGAAACAA  166
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATXN3  37   GGCGTGGGGGCCGTTGGCTCCAGACAAATAAACATGGAGTCCATCTTCCACGAGAAACAA  96

Query  167  GAAGGCTCACTTTGTGCTCAACATTGCCTGAATAACTTATTGCAAGGAGAATATTTTAGC  226
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATXN3  97   GAAGGCTCACTTTGTGCTCAACATTGCCTGAATAACTTATTGCAAGGAGAATATTTTAGC  156

Query  227  CCTGTGGAATTATCCTCAATTGCACATCAGCTGGATGAGGAGGAGAGGATGAGAATGGCA  286
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATXN3  157  CCTGTGGAATTATCCTCAATTGCACATCAGCTGGATGAGGAGGAGAGGATGAGAATGGCA  216

Query  287  GAAGGAGGAGTTACTAGTGAAGATTATCGCACGTTTTTACAGCAGCCTTCTGGAAATATG  346
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATXN3  217  GAAGGAGGAGTTACTAGTGAAGATTATCGCACGTTTTTACAGCAGCCTTCTGGAAATATG  276

Query  347  GATGACAGTGGTTTTTTCTCTATTCAGGTTATAAGCAATGCCTTGAAAGTTTGGGGTTTA  406
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATXN3  277  GATGACAGTGGTTTTTTCTCTATTCAGGTTATAAGCAATGCCTTGAAAGTTTGGGGTTTA  336

Query  407  GAACTAATCCTGTTCAACAGTCCAGAGTATCAGAGGCTCAGGATCGATCCTAT  459
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
ATXN3  337  GAACTAATCCTGTTCAACAGTCCAGAGTATCAGAGGCTCAGGATCGATCCTAT  389
```

Figure B.3 Alignment data for ATXN3 mRNA.

This insert sequence is homologous to nt 37-389 of Homo sapiens ataxin 3 (ATXN3), transcript variant o (NM_001164778.1). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for the first 107 amino acids of 154 amino acid ATXN3 protein.

Homo sapiens peptidylprolyl cis/trans isomerase, NIMA-interacting 1 (PIN1), transcript variant 1, mRNA

```
Score = 1814 bits(982),Expect = 0.0,Identities  = 990/994(99%)
Gaps = 1/994(0%), Strand = Plus/Plus

Query  110   GGAGCAGGCGCTGCGGCAGGAGGGAAGATGGCGGACGAGGAGAAGCTGCCGCCCGGCTGG  169
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   112   GGAGCAGGCGCTGCGGCAGGAGGGAAGATGGCGGACGAGGAGAAGCTGCCGCCCGGCTGG  171

Query  170   GAGAAGCGCATGAGCCGCAGCTCAGGCCGAGTGTACTACTTCAACCACATCACTAACGCC  229
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   172   GAGAAGCGCATGAGCCGCAGCTCAGGCCGAGTGTACTACTTCAACCACATCACTAACGCC  231
```

```
Query  230   AGCCAGTGGGAGCGGCCCAGCGGCAACAGCAGCAGTGGTGGCAAAAACGGGCAGGGGGAG  289
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   232   AGCCAGTGGGAGCGGCCCAGCGGCAACAGCAGCAGTGGTGGCAAAAACGGGCAGGGGGAG  291

Query  290   CCTGCCAGGGTCCGCTGCTCGCACCTGCTGGTGAAGCGCAGCCAGTCACGGCGGCCCTCG  349
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
PIN1   292   CCTGCCAGGGTCCGCTGCTCGCACCTGCTGGTGAAGCACAGCCAGTCACGGCGGCCCTCG  351

Query  350   TCCTGGCGGCAGGAGAAGATCACCCGGACCAAGGAGGAGGCCCTGGAGCTGATCAACGGC  409
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   352   TCCTGGCGGCAGGAGAAGATCACCCGGACCAAGGAGGAGGCCCTGGAGCTGATCAACGGC  411

Query  410   TACATCCAGAAGATCAAGTCGGGAGAGGAGGACTTTGAGTCTCTGGCCCCACAGTTCAGC  469
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
PIN1   412   TACATCCAGAAGATCAAGTCGGGAGAGGAGGACTTTGAGTCTCTGGCCTCACAGTTCAGC  471

Query  470   GACTGCAGCTCAGCCAAGGCCAGGGGAGACCTGGGTGCCTTCAGCAGAGGTCAGATGCAG  529
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   472   GACTGCAGCTCAGCCAAGGCCAGGGGAGACCTGGGTGCCTTCAGCAGAGGTCAGATGCAG  531

Query  530   AAGCCATTTGAAGACGCCTCGTTTGCGCTGCGGACGGGGGAGATGAGCGGGCCCGTGTTC  589
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   532   AAGCCATTTGAAGACGCCTCGTTTGCGCTGCGGACGGGGGAGATGAGCGGGCCCGTGTTC  591

Query  590   ACGGATTCCGGCATCCACATCATCCTCCGCACTGAGTGAGGGTGGGGAGCCCAGGCCTGG  649
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   592   ACGGATTCCGGCATCCACATCATCCTCCGCACTGAGTGAGGGTGGGGAGCCCAGGCCTGG  651

Query  650   CCTCGGGGCAGGGCAGGGCGGCTAGGCCGGCCAGCTCCCCCTTGCCCGCCAGCCAGTGGC  709
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   652   CCTCGGGGCAGGGCAGGGCGGCTAGGCCGGCCAGCTCCCCCTTGCCCGCCAGCCAGTGGC  711

Query  710   CGAACCCCCCACTCCCTGCCACCGTCACACAGTATTTATTGTTCCCACAATGGCTGGGAG  769
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   712   CGAACCCCCCACTCCCTGCCACCGTCACACAGTATTTATTGTTCCCACAATGGCTGGGAG  771

Query  770   GGGGCCCTTCCAGATTGGGGGCCCTGGGGTCCCCACTCCCTGTCCATCCCCAGTTGGGGC  829
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   772   GGGGCCCTTCCAGATTGGGGGCCCTGGGGTCCCCACTCCCTGTCCATCCCCAGTTGGGGC  831

Query  830   TGCGACCGCCAGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGGAGGCTCCCAGAC  889
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   832   TGCGACCGCCAGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGGAGGCTCCCAGAC  891

Query  890   CCAGGGCAGTGTGGTGGGGAGGGGTGTTCCAAAGAGAAGGCCTGGTCAGCAGAGCCGCCC  949
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
PIN1   892   CCAGGGCAGTGTGGTGGG-AGGGGTGTTCCAAAGAGAAGGCCTGGTCAGCAGAGCCGCCC  950

Query  950   CGTGTCCCCCCAGGTGCTGGAGGCAGACTCGAGGGCCGAATTGTTTCTAGTTAGGCCACG  1009
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   951   CGTGTCCCCCCAGGTGCTGGAGGCAGACTCGAGGGCCGAATTGTTTCTAGTTAGGCCACG  1010

Query  1010  CTCCTCTGTTCAGTCGCAAAGGTGAACACTCATGCGGCCCAGCCATGGGCCCTCTGAGCA  1069
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
PIN1   1011  CTCCTCTGTTCAGTCGCAAAGGTGAACACTCATGCGGCCCAGCCATGGGCCCTCTGAGCA  1070

Query  1070  ACTGTGCAGCACCCTTTCACCCNCAATTAAACCC  1103
             |||||||||||||||||||| ||||||||||||
PIN1   1071  ACTGTGCAGCACCCTTTCACCCCCAATTAAACCC  1104
```

Figure B.4 Alignment data for PIN1 mRNA.

This insert sequence is homologous to nt 112-1104 of Homo sapiens peptidylprolyl cis/trans isomerase, NIMA-interacting 1 (PIN1), transcript variant 1, mRNA (NM_006221.3). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for 161 amino acids of 163 amino acid PIN1protein.

Homo sapiens microtubule-associated protein 1B (MAP1B), mRNA

```
Score = 2098 bits(1136), Expect = 0.0, Identities = 1156/1165(99%)
Gaps = 3/1165(0%), Strand = Plus/Plus

Query  108   CATCCCCACAGACAAAACTGTCACGTACAAACACATGGACCCACCTCCAGCTCCCGTGCA  167
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  6799  CATCCCCACAGACAAAACTGTCACGTACAAACACATGGACCCACCTCCAGCTCCCGTGCA  6858

Query  168   AGACCGCAGCCCTTCGCCACGCCACCCTGATGTGTCCATGGTGGACCCAGAAGCCTTGGC  227
             |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
MAP1B  6859  AGACCGCAGCCCTTCGCCACGCCACCCTGATGTGTCCATGGTGGACCCAGAGGCCTTGGC  6918

Query  228   CATTGAGCAGAACCTGGGCAAAGCTCTAAAGAAAGATCTGAAAGAGAAGACCAAAACCAA  287
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  6919  CATTGAGCAGAACCTGGGCAAAGCTCTAAAGAAAGATCTGAAAGAGAAGACCAAAACCAA  6978

Query  288   AAAGCCAGGTACAAAGACCAAGTCATCTTCACCTGTCAAAAAGAGTGATGGGAAGTCTAA  347
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  6979  AAAGCCAGGTACAAAGACCAAGTCATCTTCACCTGTCAAAAAGAGTGATGGGAAGTCTAA  7038

Query  348   GCCCTTGGCAGCTTCACCAAAACCAGCGGGCTTGAAAGAATCCTCGGATAAAGTGTCCAG  407
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7039  GCCCTTGGCAGCTTCACCAAAACCAGCGGGCTTGAAAGAATCCTCGGATAAAGTGTCCAG  7098

Query  408   GGTGGCTTCTCCTAAGAAGAAAGAATCTGTGGAAAAGGCAGCAAAACCCACCACCACTCC  467
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7099  GGTGGCTTCTCCTAAGAAGAAAGAATCTGTGGAAAAGGCAGCAAAACCCACCACCACTCC  7158

Query  468   TGAGGTCAAAGCTGCACGTGGGGAAGAGAAAGACAAGGAGACCAAGAATGCTGCCAATGC  527
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7159  TGAGGTCAAAGCTGCACGTGGGGAAGAGAAAGACAAGGAGACCAAGAATGCTGCCAATGC  7218

Query  528   CTCTGCATCCAAGTCGGCCAAGACCGCCACTGCAGGACCAGGAACTACCAAGACGACCAA  587
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7219  CTCTGCATCCAAGTCGGCCAAGACCGCCACTGCAGGACCAGGAACTACCAAGACGACCAA  7278

Query  588   GTCATCTGCTGTGCCCCCAGGCCTCCCTGTGTATTTGGACCTGTGCTACATTCCTAACCA  647
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7279  GTCATCTGCTGTGCCCCCAGGCCTCCCTGTGTATTTGGACCTGTGCTACATTCCTAACCA  7338

Query  648   CAGCAATAGTAAGAATGTTGATGTGGAATTTTTCAAGAGAGTGCGGTCTTCCTACTACGT  707
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7339  CAGCAATAGTAAGAATGTTGATGTGGAATTTTTCAAGAGAGTGCGGTCTTCCTACTACGT  7398

Query  708   GGTGAGTGGGAATGACCCTGCTGCTGAGGAGCCCAGCCGGGCTGTCCTGGACGCTTTGTT  767
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7399  GGTGAGTGGGAATGACCCTGCTGCTGAGGAGCCCAGCCGGGCTGTCCTGGACGCTTTGTT  7458

Query  768   GGAAGGAAAGGCTCAGTGGGGCAGCAACATGCAGGTGACACTGATCCCAACTCATGACTC  827
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7459  GGAAGGAAAGGCTCAGTGGGGCAGCAACATGCAGGTGACACTGATCCCAACTCATGACTC  7518

Query  828   AGAAGTGATGAGGGAATGGTACCAGGAGACCCATGAGAAACAGCAAGATCTCAACATCAT  887
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7519  AGAAGTGATGAGGGAATGGTACCAGGAGACCCATGAGAAACAGCAAGATCTCAACATCAT  7578

Query  888   GGTTTTAGCAAGCAGCAGCACAGTGGTTATGCAAGATGAATCCTTCCCTGCATGCAAGAT  947
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7579  GGTTTTAGCAAGCAGCAGCACAGTGGTTATGCAAGATGAATCCTTCCCTGCATGCAAGAT  7638

Query  948   TGAACTGTAAAAACCAAGGCCAGCCACACCACAGGATCTGAACTTTGTTTCCAGAAATTC  1007
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7639  TGAACTGTAAAAACCAAGGCCAGCCACACCACAGGATCTGAACTTTGTTTCCAGAAATTC  7698

Query  1008  TTCAATTTGAAATCACCTTTTCTAAAAAGTCAATTCATCTAGTTAAGTCGCTGAACAATT  1067
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7699  TTCAATTTGAAATCACCTTTTCTAAAAAGTCAATTCATCTAGTTAAGTCGCTGAACAATT  7758

Query  1068  ACCTGCCAAATGCTATACTGTGTCATGGTGATGCAAGTCACTAAATTTCTCAGTTTTTGC  1127
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAP1B  7759  ACCTGCCAAATGCTATACTGTGTCATGGTGATGCAAGTCACTAAATTTCTCAGTTTTTGC  7818

Query  1128  TGATTGCTAAGGGAAATAACAGTATTTCCACAATAGGGTTCAAATTCCTGCAAA-TTACC  1186
             |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
```

```
MAP1B  7819  TGATTGCTAAGGGAAATAACAGTATTTCCACAATAGGGTTCAAATTCCTGCAAAATTACC  7878

Query  1187  TACCCCAGTTCATCTCTGCTGAACATTTGGAAACCATGCCCTAGCCAACC-AACTGATTT  1245
              |||||||||||||||||||||||||||||||||||| |||||||||| |||||| ||
MAP1B  7879  TACCCCAGTTCATCTCTGCTGAACATTTGGAAACCATGCACTAGCCAACCCAACTGACTT  7938

Query  1246  CTGGTAGGTAAAGGT-TTTGTCTTa  1269
              ||| |||||| |||  |||||||||
MAP1B 7939  CTGCTAGGTAGAGGCATTTGTCTTA  7963
```

Figure B.5 Alignment data for MAP1B mRNA.

This insert sequence is homologous to nt 6799-7963 of Homo sapiens microtubule-associated protein 1B (MAP1B), mRNA (NM_005909.3). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for the last 282 amino acids of 2468 amino acid MAP1B protein.

Homo sapiens 3-hydroxy-3-methylglutaryl-CoA synthase 1 (soluble) (HMGCS1), transcript variant 1, mRNA

```
Score = 1062 bits(575), Expect = 0.0,Identities = 575/575(100%)
Gaps = 0/575(0%),Strand = Plus/Plus

Query  106  GCGGACTGTCCTTTCGTGGCTCACTCCCTTTCCTCTGCTGCCGCTCGGTCACGCTTGTGC  165
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HMGCS1 101  GCGGACTGTCCTTTCGTGGCTCACTCCCTTTCCTCTGCTGCCGCTCGGTCACGCTTGTGC  160

Query  166  CCGAAGGAGGAAACAGTGACAGACCTGGAGACTGCAGTTCTCTATCCTTCACACAGCTCT  225
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HMGCS1 161  CCGAAGGAGGAAACAGTGACAGACCTGGAGACTGCAGTTCTCTATCCTTCACACAGCTCT  220

Query  226  TTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGA  285
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HMGCS1 221  TTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGA  280

Query  286  ATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAA  345
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HMGCS1 281  ATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAA  340

Query  346  TATGATGGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGGGCTTC  405
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HMGCS1 341  TATGATGGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGGGCTTC  400

Query  406  TGCACAGATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAG  465
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HMGCS1 401  TGCACAGATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAG  460

Query  466  AGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAATCATC  525
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HMGCS1 461  AGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAATCATC  520

Query  526  GACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACA  585
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HMGCS1 521  GACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACA  580

Query  586  GATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCTTCAAT  645
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HMGCS1 581  GATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCTTCAAT  640

Query  646  GCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGG  680
            |||||||||||||||||||||||||||||||||
HMGCS1 641  GCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGG  675
```

Figure B.6 Alignment data for HMGCS1 mRNA.

This insert sequence is homologous to nt 101-675 of 3-hydroxy-3-methylglutaryl-CoA synthase 1 soluble protein (HMGCS1) mRNA (NM_001098272.2). This partial

cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for the first 150 amino acids of 520 amino acid HMGCS1 protein.

Homo sapiens SH3-domain GRB2-like (endophilin) interacting protein 1 (SGIP1), mRNA

```
Score = 1923 bits(1041),Expect = 0.0, Identities = 1162/1215(96%)
Gaps = 31/1215(2%),Strand = Plus/Plus

Query  106   GGCAGTGCTCCCACCAGCAGTCTGGAATGCTGAACAACAGAGAATATTGTGGAAGATTCC  165
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  2428  GGCAGTGCTCCCACCAGCAGTCTGGAATGCTGAACAACAGAGAATATTGTGGAAGATTCC  2487

Query  166   TGATATCTCTCAGAAGTCAGAAAATGGAGGGGTGGGTTCTTTGTTGGCAAGATTTCAGTT  225
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  2488  TGATATCTCTCAGAAGTCAGAAAATGGAGGGGTGGGTTCTTTGTTGGCAAGATTTCAGTT  2547

Query  226   ATCTGAAGGCCCAAGCAAACCTTCTCCATTGGTTGTGCAGTTCACAAGTGAAGGAAGCAC  285
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  2548  ATCTGAAGGCCCAAGCAAACCTTCTCCATTGGTTGTGCAGTTCACAAGTGAAGGAAGCAC  2607

Query  286   CCTTTCTGGCTGTGACATTGAACTTGTTGGAGCAGGGTATCGATTTTCACTCATCAAGAA  345
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  2608  CCTTTCTGGCTGTGACATTGAACTTGTTGGAGCAGGGTATCGATTTTCACTCATCAAGAA  2667

Query  346   AAGGTTTGCTGCAGGAAAATACTTGGCAGATAACTAATGAAATCTTATGCAAGGATTTGG  405
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  2668  AAGGTTTGCTGCAGGAAAATACTTGGCAGATAACTAATGAAATCTTATGCAAGGATTTGG  2727

Query  406   AGGATTCATATAATGGAGAACTGATGTATGAGAAACAGATTTTAATTTTGGTTTGATGAA  465
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  2728  AGGATTCATATAATGGAGAACTGATGTATGAGAAACAGATTTTAATTTTGGTTTGATGAA  2787

Query  466   AACAAACCAATATCTGCACTTGGGATATATCAGGTGGAAAGTCAATGACTTTCATCTGTG  525
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  2788  AACAAACCAATATCTGCACTTGGGATATATCAGGTGGAAAGTCAATGACTTTCATCTGTG  2847

Query  526   ATTTCCCTCACACACTACCATGATGACCAGTCCTACAGTATTTACTTCTAGGTGTAATAT  585
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  2848  ATTTCCCTCACACACTACCATGATGACCAGTCCTACAGTATTTACTTCTAGGTGTAATAT  2907

Query  586   TGTTAATGGTTTTAAAATGTAATTATTGTATTTGTAAATTGTACTCTCATTCCAGTAAGG  645
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  2908  TGTTAATGGTTTTAAAATGTAATTATTGTATTTGTAAATTGTACTCTCATTCCAGTAAGG  2967

Query  646   CAGTTAGACACTTGAGTTTTAGCATTTTACCATTCCTGAAATGGATGTAATTTAAACTGT  705
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  2968  CAGTTAGACACTTGAGTTTTAGCATTTTACCATTCCTGAAATGGATGTAATTTAAACTGT  3027

Query  706   GGTATGTAAATTTAATAGTAGTATTGTTGAATGGCACAATGCTTACAGAGGTAGATTGCA  765
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  3028  GGTATGTAAATTTAATAGTAGTATTGTTGAATGGCACAATGCTTACAGAGGTAGATTGCA  3087

Query  766   TTTTGTCAATATATAAAATTTAAATATAATATTGATAGCTGTCATAAAGGGGGTGCCACA  825
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  3088  TTTTGTCAATATATAAAATTTAAATATAATATTGATAGCTGTCATAAAGGGGGTGCCACA  3147

Query  826   TATTAAAGAAACTTAAGTGGAACCAGAAGAAAAAGAAACAAACTTACTTTTCTTCAATGC  885
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  3148  TATTAAAGAAACTTAAGTGGAACCAGAAGAAAAAGAAACAAACTTACTTTTCTTCAATGC  3207

Query  886   TTAGTATGTTTTACTCTAGTGCTAAATAAAAACTCTATCTTCAAATGTTTAGTGGGTTAA  945
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SGIP1  3208  TTAGTATGTTTTACTCTAGTGCTAAATAAAAACTCTATCTTCAAATGTTTAGTGGGTTAA  3267

Query  946   ATTGAGAAACTATTTCAGAAAAAA-TTCTAAGGTTACAGCATATTCAAAGAAAAGCATTA  1004
             |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
SGIP1  3268  ATTGAGAAACTATTTCAGAAAAAAATTCTAAGGTTACAGCATATTCAAAGAAAAGCATTA  3327

Query  1005  GTTACCACTTTTTAAAAAGCttttttttCAAACTGCAATTTTCATAAAA-TGCAAACTGT  1063
             |||||||||||||||||||||||||||||||||||| |||||||||| ||||||||||
SGIP1  3328  GTTACCACTTTTTAAAAAGCTTTTTTTTCAAACTGCAAATTTCATAAAAATGCAAACTGT  3387

Query  1064  GTAAACAGGGCCTCTTATTTT-ATAACTTGTGTAAAAAGGGAAAGCA-TTCTTATTTAAA  1121
```

```
            |||||||||||||||||||| |||||||||||||||||||||||||| ||| ||||||||
SGIP1  3388 GTAAACAGGGCCTCTTATTTTTATAACTTGTGTAAAAAGGGAAAGCAATTCATATTTAAA 3447

Query  1122 GTT-AAGT-TATTTAATTTTA-TCAAAAGTAA-GAAAAAGTTGAAG-CTTA-CTACTTGC 1175
            ||| |||| |||| |||| || |||| ||||| ||| | ||||||| |||| ||||||||
SGIP1  3448 GTTTAAGTATATTAAATTATAATCAAGAGTAAAGAAGATGTTGAAGTCTTAACTACTTGC 3507

Query  1176 CCCCCT-TA-AGTTCC-CAAATG-GGGGATGG-TNAA-AATCCTCCA-A-TAAA-CCAAA 1226
            ||| || || |||| | |||||| |||||| | | || ||||   || | |||| |||||
SGIP1  3508 CCCTCTCTACAGTTTCGCAAATGTGGGGATTGCTGAATAATCAGTCAGACTAAAACCAAA 3567

Query  1227 TTTG-G-TTTTAA-ATT-CAAAA-TTTCGG-A-TTGAAAGGGTAAA-AATTTT-GCT-A- 1275
             ||| | |||||| ||| ||| | |||| | | |||||  ||| || |||||| ||| |
SGIP1  3568 ATTGTGATTTTAAGATTTCAAGACTTTCCGTAGTTGAACTGGTTAAGAATTTTTGCTTAG 3627

Query  1276 T-ACTCGGAAAAGAT 1289
            | |||| ||| ||||
SGIP1  3628 TTACTCTGAATAGAT 3642
```

Figure B.7 Alignment data for SGIP1 mRNA.

This insert sequence is homologous to nt 2428-3642 of SH3-domain GRB2-like (endophilin) interacting protein 1 (SGIP1) mRNA (NM_032291.2). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for the last 91 amino acids of 828 amino acid SGIP1 protein.

Homo sapiens nitrogen permease regulator-like 2 (S. cerevisiae) (NPRL2), mRNA

```
Score = 2041 bits(1105), Expect = 0.0,Identities = 1175/1209(97%)
Gaps = 8/1209(0%),Strand = Plus/Plus

Query  109  GAGGCTGTCTCTGACAAGTGTTCACAGGAGGTGGGGACGCCTCTGCGCGAGGAACGAGGA 168
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  317  GAGGCTGTCTCTGACAAGTGTTCACAGGAGGTGGGGACGCCTCTGCGCGAGGAACGAGGA 376

Query  169  GCTACGGGCCTGGGCCCGGTTATTGTCATGGGCAGCGGCTGCCGCATCGAATGCATATTC 228
            |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
NPRL2  377  GCTACGGGCCTGGGCCCGGTTATTGCCATGGGCAGCGGCTGCCGCATCGAATGCATATTC 436

Query  229  TTCAGCGAGTTCCACCCCACGCTGGGACCCAAGATCACCTATCAGGTCCCTGAAGACTTC 288
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  437  TTCAGCGAGTTCCACCCCACGCTGGGACCCAAGATCACCTATCAGGTCCCTGAAGACTTC 496

Query  289  ATCTCCCGAGAGCTGTTTGACACAGTCCAAGTGTACATCATCACCAAGCCAGAGCTGCAG 348
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  497  ATCTCCCGAGAGCTGTTTGACACAGTCCAAGTGTACATCATCACCAAGCCAGAGCTGCAG 556

Query  349  AACAAGCTTATCACTGTCACAGCTATGGAAAAGAAGCTGATCGGCTGTCCTGTGTGCATC 408
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  557  AACAAGCTTATCACTGTCACAGCTATGGAAAAGAAGCTGATCGGCTGTCCTGTGTGCATC 616

Query  409  GAACACAAGAAGTACAGCCGCAATGCTCTCCTCTTCAACCTGGGCTTCGTGTGTGATGCC 468
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  617  GAACACAAGAAGTACAGCCGCAATGCTCTCCTCTTCAACCTGGGCTTCGTGTGTGATGCC 676

Query  469  CAGGCCAAGACCTGCGCCCTCGAGCCCATTGTTAAAAAGCTGGCTGGCTATCTGACCACA 528
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  677  CAGGCCAAGACCTGCGCCCTCGAGCCCATTGTTAAAAAGCTGGCTGGCTATCTGACCACA 736

Query  529  CTAGAGCTAGAGAGCAGCTTCGTGTCCATGGAGGAGAGCAAGCAGAAGTTGGTGCCCATC 588
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  737  CTAGAGCTAGAGAGCAGCTTCGTGTCCATGGAGGAGAGCAAGCAGAAGTTGGTGCCCATC 796

Query  589  ATGACCATCTTGCTGGAGGAGCTAAATGCCTCAGGCCGGTGCACTCTGCCCATTGATGAG 648
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  797  ATGACCATCTTGCTGGAGGAGCTAAATGCCTCAGGCCGGTGCACTCTGCCCATTGATGAG 856

Query  649  TCCAACACCATCCACTTGAAGGTGATTGAGCAGCGGCCAGACCCTCCGGTGGCCCAGGAG 708
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  857  TCCAACACCATCCACTTGAAGGTGATTGAGCAGCGGCCAGACCCTCCGGTGGCCCAGGAG 916
```

```
Query  709   TATGATGTACCTGTCTTTACCAAAGACAAGGAGGATTTCTTCAACTCACAGTGGGACCTC  768
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  917   TATGATGTACCTGTCTTTACCAAAGACAAGGAGGATTTCTTCAACTCACAGTGGGACCTC  976

Query  769   ACTACACAACAAATCCTGCCCTACATTGATGGGTTCCGCCACATCCAGAAGATTTCAGCA  828
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  977   ACTACACAACAAATCCTGCCCTACATTGATGGGTTCCGCCACATCCAGAAGATTTCAGCA  1036

Query  829   GAGGCAGATGTGGAGCTCAACCTGGTGCGCATTGCTATCCAGAACCTGCTGTACTACGGC  888
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  1037  GAGGCAGATGTGGAGCTCAACCTGGTGCGCATTGCTATCCAGAACCTGCTGTACTACGGC  1096

Query  889   GTTGTGACACTGGTGTCCATCCTCCAGTACCCCAATGTATACTGCCCAACGCCCAAGGTC  948
             |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
NPRL2  1097  GTTGTGACACTGGTGTCCATCCTCCAGTACTCCAATGTATACTGCCCAACGCCCAAGGTC  1156

Query  949   CAGGACCTGGTAGATGACAAGTCCCTGCAAGAGGCATGTCTATCCTACGTGACCAAGCAA  1008
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
NPRL2  1157  CAGGACCTGGTAGATGACAAGTCCCTGCAAGAGGCATGTCTATCCTACGTGACCAAGCAA  1216

Query  1009  GGGCACAAGAGGGCCAGTCTCCCGGGATGTGTTCCAGCTATACTGCAGCCTGAGCCCTGG  1068
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
NPRL2  1217  GGGCACAAGAGGGCCAGTCTCC-GGGATGTGTTCCAGCTATACTGCAGCCTGAGCCCTGG  1275

Query  1069  CACTACCGTGCGAGACCTCATTGGCCGcc-cccccAGCAGCTGCAGCATGTTGATGAACG  1127
             |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
NPRL2  1276  CACTACCGTGCGAGACCTCATTGGCCGCCACCCCCAGCAGCTGCAGCATGTTGATGAACG  1335

Query  1128  GAAGCTGATCCAGTTCGGGCTTATGAAGA-CCTCATCAGGCGACTACAGAATTATCCTGG  1186
             |||||||||||||||||||||||||||| |||||||||||||||||||| |||||||
NPRL2  1336  GAAGCTGATCCAGTTCGGGCTTATGAAGAACCTCATCAGGCGACTACAGAAGTATCCTGT  1395

Query  1187  GCGGGTGACTCGGGAAAAACAGA-CCACCCTGCCCGG-TTTATACAGGCTGCCNAAGCTT  1244
             |||||||||||||||| | |||| |||||||||||||| ||||||||||||||  ||||
NPRL2  1396  GCGGGTGACTCGGGAAGAGCAGAGCCACCCTGCCCGGCTTTATACAGGCTGCCACAGCTA  1455

Query  1245  TGACAAATTTTGTTGAAAAA-AGGCTTG-GCTCCCATGAACTTGAATAACGGGTTGAAAT  1302
             |||| |  | || || || | |||| || ||| |||||| || ||  | ||| ||||||
NPRL2  1456  TGACGAGATCTGCTGCAAGACAGGCATGAGCTACCATGAGCTGGATGAGCGGCTTGAAAA  1515

Query  1303  NG-CCCCAA  1310
               | ||||||
NPRL2  1516  TGACCCCAA  1524
```

Figure B.8 Alignment data for NPRL2 mRNA.

This insert sequence is homologous to nt 317-1524 of Homo sapiens nitrogen permease regulator-like 2 (S. cerevisiae) (NPRL2), mRNA (NM_006545.4). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for the first 278 amino acids of 380 amino acid NPRL2 protein.

Homo sapiens complement component 1, q subcomponent, A chain (C1QA), mRNA

```
Score = 1626 bits(880),Expect = 0.0, Identities = 880/880(100%)
Gaps = 0/880(0%), Strand = Plus/Plus

Query  105   GGTGACCGAGGACTTGTGCCGAGCACCAGACGGGAAGAAAGGGGAGGCAGGAAGACCTGG  164
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   145   GGTGACCGAGGACTTGTGCCGAGCACCAGACGGGAAGAAAGGGGAGGCAGGAAGACCTGG  204

Query  165   CAGACGGGGGCGGCCAGGCCTCAAGGGGGAGCAAGGGGAGCCGGGGGCCCCTGGCATCCG  224
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   205   CAGACGGGGGCGGCCAGGCCTCAAGGGGGAGCAAGGGGAGCCGGGGGCCCCTGGCATCCG  264

Query  225   GACAGGCATCCAAGGCCTTAAAGGAGACCAGGGGGAACCTGGGCCCTCTGGAAACCCCGG  284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   265   GACAGGCATCCAAGGCCTTAAAGGAGACCAGGGGGAACCTGGGCCCTCTGGAAACCCCGG  324

Query  285   CAAGGTGGGCTACCCAGGGCCCAGCGGCCCCCTCGGAGCCCGTGGCATCCCGGGAATTAA  344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   325   CAAGGTGGGCTACCCAGGGCCCAGCGGCCCCCTCGGAGCCCGTGGCATCCCGGGAATTAA  384
```

```
Query  345   AGGCACCAAGGGCAGCCCAGGAAACATCAAGGACCAGCCGAGGCCAGCCTTCTCCGCCAT  404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   385   AGGCACCAAGGGCAGCCCAGGAAACATCAAGGACCAGCCGAGGCCAGCCTTCTCCGCCAT  444

Query  405   TCGGCGGAACCCCCCAATGGGGGGCAACGTGGTCATCTTCGACACGGTCATCACCAACCA  464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   445   TCGGCGGAACCCCCCAATGGGGGGCAACGTGGTCATCTTCGACACGGTCATCACCAACCA  504

Query  465   GGAAGAACCGTACCAGAACCACTCCGGCCGATTCGTCTGCACTGTACCCGGCTACTACTA  524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   505   GGAAGAACCGTACCAGAACCACTCCGGCCGATTCGTCTGCACTGTACCCGGCTACTACTA  564

Query  525   CTTCACCTTCCAGGTGCTGTCCCAGTGGGAAATCTGCCTGTCCATCGTCTCCTCCTCAAG  584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   565   CTTCACCTTCCAGGTGCTGTCCCAGTGGGAAATCTGCCTGTCCATCGTCTCCTCCTCAAG  624

Query  585   GGGCCAGGTCCGACGCTCCCTGGGCTTCTGTGACACCACCAACAAGGGGCTCTTCCAGGT  644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   625   GGGCCAGGTCCGACGCTCCCTGGGCTTCTGTGACACCACCAACAAGGGGCTCTTCCAGGT  684

Query  645   GGTGTCAGGGGGCATGGTGCTTCAGCTGCAGCAGGGTGACCAGGTCTGGGTTGAAAAAGA  704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   685   GGTGTCAGGGGGCATGGTGCTTCAGCTGCAGCAGGGTGACCAGGTCTGGGTTGAAAAAGA  744

Query  705   CCCCAAAAAGGGTCACATTTACCAGGGCTCTGAGGCCGACAGCGTCTTCAGCGGCTTCCT  764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   745   CCCCAAAAAGGGTCACATTTACCAGGGCTCTGAGGCCGACAGCGTCTTCAGCGGCTTCCT  804

Query  765   CATCTTCCCATCTGCCTGAGCCAGGGAAGGACCCCCTCCCCCACCCACCCTCTCTGGCTTC  824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   805   CATCTTCCCATCTGCCTGAGCCAGGGAAGGACCCCCTCCCCCACCCACCCTCTCTGGCTTC  864

Query  825   CATGCTCCGCCTGTAAAATGGGGGCGCTATTGCTTCAGCTGCTGAAGGGAGGGGGCTGGC  884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   865   CATGCTCCGCCTGTAAAATGGGGGCGCTATTGCTTCAGCTGCTGAAGGGAGGGGGCTGGC  924

Query  885   TCTGAGAGCCCCAGGACTGGCTGCCCCGTGACACATGCTCTAAGAAGCTCGTTTCTTAGa  944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C1QA   925   TCTGAGAGCCCCAGGACTGGCTGCCCCGTGACACATGCTCTAAGAAGCTCGTTTCTTAGA  984

Query  945   cctcttcctggaataaacatctgtgtctgtgtctgctgaa  984
             ||||||||||||||||||||||||||||||||||||||||
C1QA   985   CCTCTTCCTGGAATAAACATCTGTGTCTGTGTCTGCTGAA  1024
```

Figure B.9 Alignment data for C1QA mRNA.

This insert sequence is homologous to nt 145-1024 of Homo sapiens complement component 1, q subcomponent, A chain (C1QA), mRNA (NM_015991.2). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for the last 225 amino acids of 245 amino acid C1QA protein.

Homo sapiens MAD2 mitotic arrest deficient-like 2 (yeast) (MAD2L2), transcript variant 1, mRNA

```
Score = 1663 bits(900),Expect = 0.0,Identities = 914/921(99%)
Gaps = 3/921(0%), Strand = Plus/Plus

Query  153   GTGGTGGCCGATGTGCTCTGCGAGTTCCTGGAGGTGGCTGTGCATCTCATCCTCTACGTG  212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 261   GTGGTGGCCGATGTGCTCTGCGAGTTCCTGGAGGTGGCTGTGCATCTCATCCTCTACGTG  320

Query  213   CGCGAGGTCTACCCCGTGGGCATCTTCCAGAAACGCAAGAAGTACAACGTGCCGGTCCAG  272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 321   CGCGAGGTCTACCCCGTGGGCATCTTCCAGAAACGCAAGAAGTACAACGTGCCGGTCCAG  380

Query  273   ATGTCCTGCCACCCGGAGCTGAATCAGTATATCCAGGACACGCTGCACTGCGTCAAGCCA  332
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 381   ATGTCCTGCCACCCGGAGCTGAATCAGTATATCCAGGACACGCTGCACTGCGTCAAGCCA  440

Query  333   CTCCTGGAGAAGAATGATGTGGAGAAAGTGGTGGTGGTGATTTTGGATAAAGAGCACCGC  392
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 441      CTCCTGGAGAAGAATGATGTGGAGAAAGTGGTGGTGGTGATTTTGGATAAAGAGCACCGC  500

Query  393      CCAGTGGAGAAATTCGTCTTTGAGATCACCCAGCCTCCACTGCTGTCCATCAGCTCAGAC  452
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 501      CCAGTGGAGAAATTCGTCTTTGAGATCACCCAGCCTCCACTGCTGTCCATCAGCTCAGAC  560

Query  453      TCGCTGTTGTCTCATGTGGAGCAGCTGCTCCGGGCCTTCATCCTGAAGATCAGCGTGTGC  512
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 561      TCGCTGTTGTCTCATGTGGAGCAGCTGCTCCGGGCCTTCATCCTGAAGATCAGCGTGTGC  620

Query  513      GATGCCGTCCTGGACCACAACCCCCCAGGCTGTACCTTCACAGTCCTGGTGCACACGAGA  572
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 621      GATGCCGTCCTGGACCACAACCCCCCAGGCTGTACCTTCACAGTCCTGGTGCACACGAGA  680

Query  573      GAAGCCGCCACTCGCAACATGGAGAAGATCCAGGTCATCAAGGATTTCCCCTGGATCCTG  632
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 681      GAAGCCGCCACTCGCAACATGGAGAAGATCCAGGTCATCAAGGATTTCCCCTGGATCCTG  740

Query  633      GCGGATGAGCAGGATGTCCACATGCATGACCCCCGGCTGATACCACTAAAAACCATGACG  692
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 741      GCGGATGAGCAGGATGTCCACATGCATGACCCCCGGCTGATACCACTAAAAACCATGACG  800

Query  693      TCGGACATTTTAAAGATGCAGCTTTGCGTGGAAGAGCGCGCTCATAAAGGCAGCTGAGGG  752
                |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
MAD2L2 801      TCGGACATTTTAAAGATGCAGCTTTACGTGGAAGAGCGCGCTCATAAAGGCAGCTGAGGG  860

Query  753      GGCACCTGCCACCCCACTGATGCCCAAACTGTCAGACTTTGGGGGATCCCCGCCTAGGGC  812
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 861      GGCACCTGCCACCCCACTGATGCCCAAACTGTCAGACTTTGGGGGATCCCCGCCTAGGGC  920

Query  813      AGTGCTGCATGGCTGCCCTGATTCCAAGTGCTCTTATCGCCTCTGNTGTGTGGATCGCCC  872
                |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
MAD2L2 921      AGTGCTGCATGGCTGCCCTGATTCCAAGTGCTCTTATCGCCTCTG-TGTGTGGATCGCCC  979

Query  873      GCCCCAGCCCGGGGCCGCTCAGGTCTGCTTGGAGGATGCCTCCCCCAGGAGGGCAGTGAG  932
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 980      GCCCCAGCCCGGGGCCGCTCAGGTCTGCTTGGAGGATGCCTCCCCCAGGAGGGCAGTGAG  1039

Query  933      GGATGCCGCAACCTCGACTTCTCAGCCTCCTGGGGTTCCGCCGGCCAACACTGTCTGTCT  992
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MAD2L2 1040     GGATGCCGCAACCTCGACTTCTCAGCCTCCTGGGGTTCCGCCGGCCAACACTGTCTGTCT  1099

Query  993      CAA-TACTGTGGCTGTGAGTTGTTTCAATAAAGGGGCCCCAAGGGCTGGGCTGAGaaaaa  1051
                ||| ||||||| ||||||||||||||||||||||||||||||||||||||||||| |||||
MAD2L2 1100     CAAATACTGTG-CTGTGAGTTGTTTCAATAAAGGGGCCCCAAGGGCTGGGCTGAAAAAAA  1158

Query  1052     aaaaaaaannaaaaaaaaaa  1072
                ||||||||  ||||||||||
MAD2L2 1159     AAAAAAAAAAAAAAAAAAAA  1179
```

Figure B.10 Alignment data for MAD2L2 mRNA.

This insert sequence is homologous to nt 261-1179 of Homo sapiens MAD2L2 mitotic arrest deficient-like 2 (yeast) (MAD2L2), transcript variant 1, mRNA (NM_001127325.1). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for the last 198 amino acids of 211 amino acid MAD2B protein, (protein product of MAD2L2).

<u>Homo sapiens syndecan binding protein (syntenin) (SDCBP), transcript variant 1, mRNA</u>

```
Score = 2045 bits(1107),Expect = 0.0, Identities = 1223/1281(95%)
Gaps = 12/1281(0%),Strand = Plus/Plus

Query  107      TACATGGGGCTGAGTTTAAATGAAGAAGAAATACGTGCAAATGTGGCCGTGGTTTCTGGT  166
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  316      TACATGGGGCTGAGTTTAAATGAAGAAGAAATACGTGCAAATGTGGCCGTGGTTTCTGGT  375
```

```
Query  167   GCACCACTTCAGGGGCAGTTGGTAGCAAGACCTTCCAGTATAAACTATATGGTGGCTCCT  226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  376   GCACCACTTCAGGGGCAGTTGGTAGCAAGACCTTCCAGTATAAACTATATGGTGGCTCCT  435

Query  227   GTAACTGGTAATGATGTTGGAATTCGTAGAGCAGAAATTAAGCAAGGGATTCGTGAAGTC  286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  436   GTAACTGGTAATGATGTTGGAATTCGTAGAGCAGAAATTAAGCAAGGGATTCGTGAAGTC  495

Query  287   ATTTTGTGTAAGGATCAAGATGGAAAAATTGGACTCAGGCTTAAATCAATAGATAATGGT  346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  496   ATTTTGTGTAAGGATCAAGATGGAAAAATTGGACTCAGGCTTAAATCAATAGATAATGGT  555

Query  347   ATATTTGTTCAGCTAGTCCAGGCTAATTCTCCAGCCTCATTGGTTGGTCTGAGATTTGGG  406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  556   ATATTTGTTCAGCTAGTCCAGGCTAATTCTCCAGCCTCATTGGTTGGTCTGAGATTTGGG  615

Query  407   GACCAAGTACTTCAGATCAATGGTGAAAACTGTGCAGGATGGAGCTCTGATAAAGCGCAC  466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  616   GACCAAGTACTTCAGATCAATGGTGAAAACTGTGCAGGATGGAGCTCTGATAAAGCGCAC  675

Query  467   AAGGTGCTCAAACAGGCTTTTGGAGAGAAGATTACCATGACCATTCGTGACAGGCCCTTT  526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  676   AAGGTGCTCAAACAGGCTTTTGGAGAGAAGATTACCATGACCATTCGTGACAGGCCCTTT  735

Query  527   GAACGGACGATTACCATGCATAAGGATAGCACTGGACATGTTGGTTTTATCTTTAAAAAT  586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  736   GAACGGACGATTACCATGCATAAGGATAGCACTGGACATGTTGGTTTTATCTTTAAAAAT  795

Query  587   GGAAAAATAACATCCATAGTGAAAGATAGCTCTGCAGCCAGAAATGGTCTTCTCACGGAA  646
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  796   GGAAAAATAACATCCATAGTGAAAGATAGCTCTGCAGCCAGAAATGGTCTTCTCACGGAA  855

Query  647   CATAACATCTGTGAAATCAATGGACAGAATGTCATTGGATTGAAGGACTCTCAAATTGCA  706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  856   CATAACATCTGTGAAATCAATGGACAGAATGTCATTGGATTGAAGGACTCTCAAATTGCA  915

Query  707   GACATACTGTCAACATCTGGGACTGTAGTTACTATTACAATCATGCCTGCTTTTATCTTT  766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  916   GACATACTGTCAACATCTGGGACTGTAGTTACTATTACAATCATGCCTGCTTTTATCTTT  975

Query  767   GAACATATTATTAAGCGGATGGCACCAAGCATTATGAAAAGCCTAATGGACCACACCATT  826
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  976   GAACATATTATTAAGCGGATGGCACCAAGCATTATGAAAAGCCTAATGGACCACACCATT  1035

Query  827   CCTGAGGTTTAAAATTCACGGCACCATGGAAATGTAGCTGAACGTCTCCAGTTTCCTTCT  886
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  1036  CCTGAGGTTTAAAATTCACGGCACCATGGAAATGTAGCTGAACGTCTCCAGTTTCCTTCT  1095

Query  887   TTGGCAACTTCTGTATTATGCACGTGAAGCCTTCCCGGAGCCAGCGAGCATATGCTGCAT  946
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  1096  TTGGCAACTTCTGTATTATGCACGTGAAGCCTTCCCGGAGCCAGCGAGCATATGCTGCAT  1155

Query  947   GAGGACCTTTCTATCTTACATTATGGCTGGGAATCTTACTCTTTCATCTGATACCTTGTT  1006
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  1156  GAGGACCTTTCTATCTTACATTATGGCTGGGAATCTTACTCTTTCATCTGATACCTTGTT  1215

Query  1007  CAGATTTCAAAATAGTTGTAGCCTTATCCTGGTTTTACAGATGTGAAACTTTCAAGAGAT  1066
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SDCBP  1216  CAGATTTCAAAATAGTTGTAGCCTTATCCTGGTTTTACAGATGTGAAACTTTCAAGAGAT  1275

Query  1067  TTACTGACTTTCCTAAAATAGTTTCTCTACTGGAAACCTGATGCTTTTATAAGCCATTGT  1126
             |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
SDCBP  1276  TTACTGACTTTCCTAGAATAGTTTCTCTACTGGAAACCTGATGCTTTTATAAGCCATTGT  1335

Query  1127  GATTAAGATGACTGTTACAGGCTTTACTTTTGTGGGAAAACCAGTCACCTTTTTCCTAAG  1186
             ||||| ||||||||||||||||  |||| ||| |||||||||||||||| ||||| |
SDCBP  1336  GATTAGGATGACTGTTACAGGCTTAGCTTT-GTGTGAAAACCAGTCACCTTTCTCCTAGG  1394

Query  1187  GAATGAAGAAGGGCGGTCNAAATTACTTTAAGTTCTTAANCAAACTTGGATCTTTAACAN  1246
              ||||| | || || ||  | ||||||||| |||||  | || ||||| ||||||||||
SDCBP  1395  TAATGA-GTAGTGCTGTTCATATTACTTTA-GTTCTATAGCATACTTGCATCTTTAACAT  1452

Query  1247  NGCTATCA-AGGACATTTAAAANGATTGCCTTTGAAtttttttt-AAATCTCGGGG-GTG  1303
              ||||||| || |||||||| || |||||||||||| |||||||| |||| || | | |||
SDCBP  1453  -GCTATCATAGTACATTTAGAATGATTGCCTTTGATTTTTTTTTTAAAT-TCTGTGTGTG  1510

Query  1304  GGTNGGGGGAAAANGCCANTaaaaaa-aCAGGGTTCTTTCCCTGGAGGC-TTAAACAAGG  1361
```

```
             || | |  |||| |||| | || || || || ||| |||| || | || |||||||  |
SDCBP  1511  TGT-GTGT-AAAATGCCAATTAAGAACACTGGTTTCATTCCATGTAAGCATTAAACAGTG  1568

Query  1362  GATTGGGGTTTCAAGAAATTG  1382
              ||   |||||||||| ||||
SDCBP  1569  TATGTAGGTTTCAAGAGATTG  1589
```

Figure B.11 Alignment data for SDCBP mRNA.

This insert sequence is homologous to nt 316-1589 of Homo sapiens syndecan binding protein (syntenin) (SDCBP), transcript variant 1, mRNA (NM_001007067.1). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for the last 246 amino acids of 298 amino acid SDCBP protein.

Homo sapiens SEC31 homolog A (S. cerevisiae) (SEC31A), transcript variant 6, mRNA

```
Score = 1892 bits(1024),Expect = 0.0,Identities = 1033/1037(99%)
Gaps = 2/1037(0%),Strand = Plus/Plus

Query  107   CAACCAGCCAAATATCATGCAGCTTCGTGACAGACTTTGTAGAGCACAAGGAGAGCCTGT  166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2338  CAACCAGCCAAATATCATGCAGCTTCGTGACAGACTTTGTAGAGCACAAGGAGAGCCTGT  2397

Query  167   AGCAGGACATGAATCACCTAAAATTCCGTACGAGAAACAGCAGCTCCCCAAGGGCAGGCC  226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2398  AGCAGGACATGAATCACCTAAAATTCCGTACGAGAAACAGCAGCTCCCCAAGGGCAGGCC  2457

Query  227   TGGACCAGTTGCTGGCCACCACCAGATGCCAAGAGTTCAAACTCAACAATATTATCCCCA  286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2458  TGGACCAGTTGCTGGCCACCACCAGATGCCAAGAGTTCAAACTCAACAATATTATCCCCA  2517

Query  287   TGGAGAAAATCCTCCACCTCCGGGTTTCATAATGCATGGAAATGTTAATCCAAATGCTGC  346
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2518  TGGAGAAAATCCTCCACCTCCGGGTTTCATAATGCATGGAAATGTTAATCCAAATGCTGC  2577

Query  347   TGGTCAGCTTCCCACATCTCCAGGTCATATGCACACCCAGGTACCACCTTATCCACAGCC  406
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2578  TGGTCAGCTTCCCACATCTCCAGGTCATATGCACACCCAGGTACCACCTTATCCACAGCC  2637

Query  407   ACAGCCTTATCAACCAGCCCAGCCGTATCCCTTCGGAACAGGGGGGTCAGCAATGTATCG  466
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2638  ACAGCCTTATCAACCAGCCCAGCCGTATCCCTTCGGAACAGGGGGGTCAGCAATGTATCG  2697

Query  467   ACCTCAGCAGCCTGTTGCTCCTCCTACTTCAAACGCTTACCCTAACACCCCTTACATATC  526
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2698  ACCTCAGCAGCCTGTTGCTCCTCCTACTTCAAACGCTTACCCTAACACCCCTTACATATC  2757

Query  527   TTCTGCTTCTTCCTATACTGGGCAGTCTCAGCTGTACGCAGCACAGCACCAGGCCTCTTC  586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2758  TTCTGCTTCTTCCTATACTGGGCAGTCTCAGCTGTACGCAGCACAGCACCAGGCCTCTTC  2817

Query  587   ACCTACCTCCAGCCCTGCTACTTCTTTCCCTCCTCCCCCTTCCTCTGGAGCATCCTTCCA  646
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2818  ACCTACCTCCAGCCCTGCTACTTCTTTCCCTCCTCCCCCTTCCTCTGGAGCATCCTTCCA  2877

Query  647   GCATGGCGGACCAGGAGCTCCACCATCATCTTCAGCTTATGCACTGCCTCCTGGAACAAC  706
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2878  GCATGGCGGACCAGGAGCTCCACCATCATCTTCAGCTTATGCACTGCCTCCTGGAACAAC  2937

Query  707   AGGTCCTCAGAATGGTTGGAATGACCCTCCAGCTTTGAACAGAGTACCCAAAAAGAAGAA  766
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2938  AGGTCCTCAGAATGGTTGGAATGACCCTCCAGCTTTGAACAGAGTACCCAAAAAGAAGAA  2997

Query  767   GATGCCTGAAAACTTCATGCCTCCTGTTCCCATCACATCACCAATCATGAACCCGTTGGG  826
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 2998  GATGCCTGAAAACTTCATGCCTCCTGTTCCCATCACATCACCAATCATGAACCCGTTGGG  3057

Query  827   TGACCCCCAGTCACAAATGCTGCAGCAACAGCCTTCAGCTCCAGTACCACTGTCAAGCCA  886
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 3058  TGACCCCCAGTCACAAATGCTGCAGCAACAGCCTTCAGCTCCAGTACCACTGTCAAGCCA  3117
```

```
Query  887   GTCTTCATTCCCACAGCCACATCTTCCAGGTGGCCAGCCCTTCCATGGCGTACAGCAACC  946
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 3118  GTCTTCATTCCCACAGCCACATCTTCCAGGTGGCCAGCCCTTCCATGGCGTACAGCAACC  3177

Query  947   TCTTGGTCAAACAGGCATGCCACCATCTTTTTCAAAGCCCAATATTGAAGGTGCCCCAGG  1006
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SEC31A 3178  TCTTGGTCAAACAGGCATGCCACCATCTTTTTCAAAGCCCAATATTGAAGGTGCCCCAGG  3237

Query  1007  GGCTCCTATTGGAAATACCTTCCAGCATGTGCAGTCTTTGCCAACaaaaaaa-TTACCAA  1065
             |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
SEC31A 3238  GGCTCCTATTGGAAATACCTTCCAGCATGTGCAGTCTTTGCCAACAAAAAAAATTACCAA  3297

Query  1066  GAAACCTATTCCAGATGAGCACCTCATTCTAAAGACCCCTTTTGAGGATCTTATTCAGCG  1125
             |||||||||||||||||||||||||||||||||| | ||||||||||||||||||||
SEC31A 3298  GAAACCTATTCCAGATGAGCACCTCATTCTAAAGACCACATTTGAGGATCTTATTCAGCG  3357

Query  1126  CTGCCTT-CTTCAGCAA  1141
             ||||||| |||||||||
SEC31A 3358  CTGCCTTTCTTCAGCAA  3374
```

Figure B.12 Alignment data for SEC31A mRNA.

This insert sequence is homologous to nt 2338-3374 of Homo sapiens SEC31 homolog A (S. cerevisiae) (SEC31A), transcript variant 6, mRNA (NM_001191049.1). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD and codes for an intermediate 330 amino acids of 1200 amino acid SEC31A protein.

Homo sapiens chromodomain helicase DNA binding protein 3 (CHD3), transcript variant 1, mRNA

```
Score = 1574 bits(852), Expect = 0.0, Identities = 857/859(99%)
Gaps  = 2/859(0%),Strand = Plus/Plus

Query  104   GAGGTGGAGCCTCCCCAGCCGTTTCCCTGCAGAATCAGCTCTGTCTCATGTGGAAGTGGA  163
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3   6480  GAGGTGGAGCCTCCCCAGCCGTTTCCCTGCAGAATCAGCTCTGTCTCATGTGGAAGTGGA  6539

Query  164   GAATCAGCCTTGCCTGGCCTTTAGGAACTTTTGTGGGGAAGAGAGCTTTGAagagaggag  223
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3   6540  GAATCAGCCTTGCCTGGCCTTTAGGAACTTTTGTGGGGAAGAGAGCTTTGAAGAGAGGAG  6599

Query  224   ggggactttagagagggatgaaaatgagccctgggagggaggaagggacgaggaggggTG  283
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3   6600  GGGGACTTTAGAGAGGGATGAAAATGAGCCCTGGGAGGGAGGAAGGGACGAGGAGGGGTG  6659

Query  284   GCTGCATGTTACCGTCCCCTACCTCTCCCCACGTGGAGGGTGGAGCAGTTATGAGGGAGG  343
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3   6660  GCTGCATGTTACCGTCCCCTACCTCTCCCCACGTGGAGGGTGGAGCAGTTATGAGGGAGG  6719

Query  344   AAGTCAACTGCTGTTCAGCCTCAGAATAAAGGTGCCGTTCACTGGCTCAGTTACCTCCTG  403
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3   6720  AAGTCAACTGCTGTTCAGCCTCAGAATAAAGGTGCCGTTCACTGGCTCAGTTACCTCCTG  6779

Query  404   TGTACCGGCATCTTGTGTTGGGAATGTTCCCCCCTCCCTAGGGACCAAGGACCACCCCTA  463
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3   6780  TGTACCGGCATCTTGTGTTGGGAATGTTCCCCCCTCCCTAGGGACCAAGGACCACCCCTA  6839

Query  464   CAAAAAGAGTAATGGTTGGGTGATACTCCCTCAAGCCAAAGAGGAGCTCCCCAACCTGTT  523
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3   6840  CAAAAAGAGTAATGGTTGGGTGATACTCCCTCAAGCCAAAGAGGAGCTCCCCAACCTGTT  6899

Query  524   CTAGGGACCCAGGTAACCTAGAAGGGTGGGAGAGAATACAATGGGCCAGATGTGGTGGAA  583
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3   6900  CTAGGGACCCAGGTAACCTAGAAGGGTGGGAGAGAATACAATGGGCCAGATGTGGTGGAA  6959

Query  584   GCCCAGCTCTGGGGCTCAGGTTCCTGGAAGACTTCTACTACCCTCCCTCCTCAAGGCCTG  643
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3     6960   GCCCAGCTCTGGGGCTCAGGTTCCTGGAAGACTTCTACTACCCTCCCTCCTCAAGGCCTG  7019

Query    644    GATACAGACTAAATTTGTATAAGTCAGGCAGGGGACCTAGTCAGGGTCTTGGGAGCTACC  703
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3     7020   GATACAGACTAAATTTGTATAAGTCAGGCAGGGGACCTAGTCAGGGTCTTGGGAGCTACC  7079

Query    704    TTGTCGTTGGGACCAGAGCAAAATAGTGGAGGGCAGGCTAGGGAAATGTGGGCACATCCC  763
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3     7080   TTGTCGTTGGGACCAGAGCAAAATAGTGGAGGGCAGGCTAGGGAAATGTGGGCACATCCC  7139

Query    764    CCCTCCCAGGAGGGGCCGGGGAGAGTGGCAGTTTGCATGGCGAACCCCCCACTTCCTCTT  823
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3     7140   CCCTCCCAGGAGGGGCCGGGGAGAGTGGCAGTTTGCATGGCGAACCCCCCACTTCCTCTT  7199

Query    824    TGCTGCCCCTTCACTTTCTTGCTGCCCCTTTCCCAGTCTCTCTTCACACCCACTCCTGGT  883
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CHD3     7200   TGCTGCCCCTTCACTTTCTTGCTGCCCCTTTCCCAGTCTCTCTTCACACCCACTCCTGGT  7259

Query    884    CTGTCCTGATCCCCTCTTCTGTATCAGGGTTTATTGGTTGTACATATAAATTATACTttc  943
                |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
CHD3     7260   CTGTCCTGATCCCCTCTTCTGTATCAGG-TTTATTGGTTGTACATATAAATTATACTTTC  7318

Query    944    ctttct-aaaaaaaaaaaa  961
                ||||||  ||||||||||||
CHD3     7319   CTTTCTGAAAAAAAAAAAA  7337
```

Figure B.13 Alignment data for CHD3 mRNA.

This insert sequence is homologous to nt 6480-7337 of Homo sapiens chromodomain helicase DNA binding protein 3 (CHD3), transcript variant 1, mRNA (NM_001005273.2). This partial cDNA sequence is fused in frame to the GAL4 DNA-BD. The peptide sequence coded by this transcript does not correspond to a peptide inside CHD3 protein sequence.

Homo sapiens major facilitator superfamily domain containing 2A (MFSD2A), transcript variant 1, mRNA

```
Score = 1836 bits(994), Expect = 0.0,Identities = 1097/1145(96%)
Gaps = 17/1145(1%),Strand = Plus/Plus

Query    106    GGCTGGTCATGAGCCACGGCCCATACATCAAACTTATTACTGGCTTCCTCTTCACCTCCT  165
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A   1077   GGCTGGTCATGAGCCACGGCCCATACATCAAACTTATTACTGGCTTCCTCTTCACCTCCT  1136

Query    166    TGGCTTTCATGCTGGTGGAGGGGAACTTTGTCTTGTTTTGCACCTACACCTTGGGCTTCC  225
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A   1137   TGGCTTTCATGCTGGTGGAGGGGAACTTTGTCTTGTTTTGCACCTACACCTTGGGCTTCC  1196

Query    226    GCAATGAATTCCAGAATCTACTCCTGGCCATCATGCTCTCGGCCACTTTAACCATTCCCA  285
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A   1197   GCAATGAATTCCAGAATCTACTCCTGGCCATCATGCTCTCGGCCACTTTAACCATTCCCA  1256

Query    286    TCTGGCAGTGGTTCTTGACCCGGTTTGGCAAGAAGACAGCTGTATATGTTGGGATCTCAT  345
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A   1257   TCTGGCAGTGGTTCTTGACCCGGTTTGGCAAGAAGACAGCTGTATATGTTGGGATCTCAT  1316

Query    346    CAGCAGTGCCATTTCTCATCTTGGTGGCCCTCATGGAGAGTAACCTCATCATTACATATG  405
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A   1317   CAGCAGTGCCATTTCTCATCTTGGTGGCCCTCATGGAGAGTAACCTCATCATTACATATG  1376

Query    406    CGGTAGCTGTGGCAGCTGGCATCAGTGTGGCAGCTGCCTTCTTACTACCCTGGTCCATGC  465
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A   1377   CGGTAGCTGTGGCAGCTGGCATCAGTGTGGCAGCTGCCTTCTTACTACCCTGGTCCATGC  1436

Query    466    TGCCTGATGTCATTGACGACTTCCATCTGAAGCAGCCCCACTTCCATGGAACCGAGCCCA  525
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A   1437   TGCCTGATGTCATTGACGACTTCCATCTGAAGCAGCCCCACTTCCATGGAACCGAGCCCA  1496
```

```
Query  526   TCTTCTTCTCCTTCTATGTCTTCTTCACCAAGTTTGCCTCTGGAGTGTCACTGGGCATTT  585
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A 1497  TCTTCTTCTCCTTCTATGTCTTCTTCACCAAGTTTGCCTCTGGAGTGTCACTGGGCATTT  1556

Query  586   CTACCCTCAGTCTGGACTTTGCAGGGTACCAGACCCGTGGCTGCTCGCAGCCGGAACGTG  645
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A 1557  CTACCCTCAGTCTGGACTTTGCAGGGTACCAGACCCGTGGCTGCTCGCAGCCGGAACGTG  1616

Query  646   TCAAGTTTACACTGAACATGCTCGTGACCATGGCTCCCATAGTTCTCATCCTGCTGGGCC  705
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A 1617  TCAAGTTTACACTGAACATGCTCGTGACCATGGCTCCCATAGTTCTCATCCTGCTGGGCC  1676

Query  706   TGCTGCTCTTCAAAATGTACCCCATTGATGAAGAGAGGCGGCGGCAGAATAAGAAGGCCC  765
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
MFSD2A 1677  TGCTGCTCTTCAAAATGTACCCCATTGATGAGGAGAGGCGGCGGCAGAATAAGAAGGCCC  1736

Query  766   TGCAGGCACTGAGGGACGAGGCCAGCAGCTCTGGCTGCTCAGAAACAGACTCCACAGAGC  825
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A 1737  TGCAGGCACTGAGGGACGAGGCCAGCAGCTCTGGCTGCTCAGAAACAGACTCCACAGAGC  1796

Query  826   TGGCTAGCATCCTCTAGGGCCCGCCACGTTGCCCGAAGCCACCATGCAGAAGGCCACAGA  885
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A 1797  TGGCTAGCATCCTCTAGGGCCCGCCACGTTGCCCGAAGCCACCATGCAGAAGGCCACAGA  1856

Query  886   AGGGATCAGGACCTGTCTGCCGGCTTGCTGAGCAGCTGGACTGCAGGTGCTAGGAAGGGA  945
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A 1857  AGGGATCAGGACCTGTCTGCCGGCTTGCTGAGCAGCTGGACTGCAGGTGCTAGGAAGGGA  1916

Query  946   ACTGAAGACTCAAGGAGGTGGCCCAGGACACTTGCTGTGCTCActgtggggccggctgct  1005
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MFSD2A 1917  ACTGAAGACTCAAGGAGGTGGCCCAGGACACTTGCTGTGCTCACTGTGGGGCCGGCTGCT  1976

Query  1006  ctggggcctcctgcctccc-tctgcctgcctgggggccaa-ccctggggctgccactgg  1063
             ||| ||||||||||||||| |||||||||||| |||||||| ||||||||||||||||||
MFSD2A 1977  CTGTGGCCTCCTGCCTCCCCTCTGCCTGCCTGTGGGGCCAAGCCCTGGGGCTGCCACTGT  2036

Query  1064  gAAAATGCCAAGGAATGATCGGGCCTACCCCGAAA-ACTAAAGGAAAAAACttttttttt  1122
             ||| |||||||||| |||||||||||| |||| || ||||| | | ||| ||||||||||
MFSD2A 2037  GAATATGCCAAGGACTGATCGGGCCTAGCCCGGAACACTAATGTAGAAACCTTTTTTTTT  2096

Query  1123  ACAAAGCCCAATAAA-AACTTAAAGACTGGG-ACAANNCAAAGGGGGGTGTTGGTAAAT-  1179
             ||| |||| ||| || ||||||| ||||| | |||   ||| | | | |||  ||| ||
MFSD2A 2097  ACAGAGCCTAATTAATAACTTAATGACTGTGTACATAGCAATGTGTG-TGTATGTATATG  2155

Query  1180  TCTGGGACGC-ATTAaagggtttaattttc-taaaa-----aaa--aaaaaaaaaaaaaa  1230
             |||| || || ||||| |   ||||||||| |||||     |||  ||||||||||||||
MFSD2A 2156  TCTGTGA-GCTATTAATGTTATTAATTTTCATAAAAGCTGGAAAGCAAAAAAAAAAAAAA  2214

Query  1231  aaaaa  1235
             |||||
MFSD2A 2215  AAAAA  2219
```

Figure B.14 Alignment data for MFSD2A mRNA.

This insert sequence is homologous to nt 1077-2219 of major facilitator superfamily domain-containing protein 2A (MFSD2A) mRNA (NM_001136493.1). This partial cDNA sequence is not fused in frame to the GAL4 DNA-BD and because of this, it does not code a part or full of MFSD2A protein.

<u>Homo sapiens guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1 (GNB2L1), mRNA</u>

```
Score = 1531 bits(829), Expect  = 0.0, Identities = 843/849(99%)
Gaps = 3/849(0%), Strand = Plus/Plus

Query  105   GGGATGAGACCAACTATGGAATTCCACAGCGTGGCTCTGCGGGGTCACTCCCACTTTGTT  164
             ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
GNB2L1 246   GGGATGAGACCAACTATGGAATTCCACAGCGT-GCTCTGCGGGGTCACTCCCACTTTGTT  304

Query  165   AGTGATGTGGTTATCTCCTCAGATGGCCAGTTTGCCCTCTCAGGCTCCTGGGATGGAACC  224
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
GNB2L1 305   AGTGATGTGGTTATCTCCTCAGATGGCCAGTTTGCCCTCTCAGGCTCCTGGGATGGAACC  364

Query  225   CTGCGCCTCTGGGATCTCACAACGGGCACCACCACGAGGCGATTTGTGGGCCATACCAAG  284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GNB2L1 365   CTGCGCCTCTGGGATCTCACAACGGGCACCACCACGAGGCGATTTGTGGGCCATACCAAG  424

Query  285   GATGTGCTGAGTGTGGCCTTCTCCTCTGACAACCGGCAGATTGTCTCTGGATCTCGAGAT  344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GNB2L1 425   GATGTGCTGAGTGTGGCCTTCTCCTCTGACAACCGGCAGATTGTCTCTGGATCTCGAGAT  484

Query  345   AAAACCATCAAGCTATGGAATACCCTGGGTGTGTGCAAATACACTGTCCAGGATGAGAGC  404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GNB2L1 485   AAAACCATCAAGCTATGGAATACCCTGGGTGTGTGCAAATACACTGTCCAGGATGAGAGC  544

Query  405   CACTCAGAGTGGGTGTCTTGTGTCCGCTTCTCGCCCAACAGCAGCAACCCTATCATCGTC  464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GNB2L1 545   CACTCAGAGTGGGTGTCTTGTGTCCGCTTCTCGCCCAACAGCAGCAACCCTATCATCGTC  604

Query  465   TCCTGTGGCTGGGACAAGCTGGTCAAGGTATGGAACCTGGCTAACTGCAAGCTGAAGACC  524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GNB2L1 605   TCCTGTGGCTGGGACAAGCTGGTCAAGGTATGGAACCTGGCTAACTGCAAGCTGAAGACC  664

Query  525   AACCACATTGGCCACACAGGCTATCTGAACACGGTGACTGTCTCTCCGGATGGATCCCTC  584
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
GNB2L1 665   AACCACATTGGCCACACAGGCTATCTGAACACGGTGACTGTCTCTCCAGATGGATCCCTC  724

Query  585   TGTGCTTCTGGAGGCAAGGATGGCCAGGCCATGTTATGGGATCTCAACGAAGGCAAACAC  644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GNB2L1 725   TGTGCTTCTGGAGGCAAGGATGGCCAGGCCATGTTATGGGATCTCAACGAAGGCAAACAC  784

Query  645   CTTTACACGCTAGATGGTGGGGACATCATCAACGCCCTGTGCTTCAGCCCTAACCGCTAC  704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GNB2L1 785   CTTTACACGCTAGATGGTGGGGACATCATCAACGCCCTGTGCTTCAGCCCTAACCGCTAC  844

Query  705   TGGCTGTGTGCTGCCACAGGCCCCAGCATCAAGATCTGGGATTTAGAGGGAAAGATCATT  764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GNB2L1 845   TGGCTGTGTGCTGCCACAGGCCCCAGCATCAAGATCTGGGATTTAGAGGGAAAGATCATT  904

Query  765   GTAGATGAACTGAAGCGAGAAGTTATCAGTACCAGCAGCAAGGCAGAACCACCCCAGTGC  824
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
GNB2L1 905   GTAGATGAACTGAAGCAAGAAGTTATCAGTACCAGCAGCAAGGCAGAACCACCCCAGTGC  964

Query  825   ACCTCCCTGGCCTGGTCTGCTGATGGCCAGACTCTGTTTGCTGGCTACACGGACAACCTG  884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GNB2L1 965   ACCTCCCTGGCCTGGTCTGCTGATGGCCAGACTCTGTTTGCTGGCTACACGGACAACCTG  1024

Query  885   GTGCGAGTGTGGCAGGTGACCATTGGCACACGCTAGAAGTTTATGGCAGAGCTTTTNCaa  944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||
GNB2L1 1025  GTGCGAGTGTGGCAGGTGACCATTGGCACACGCTAGAAGTTTATGGCAGAGCTTTA-CAA  1083

Query  945   a-aaaaaaa  952
              | |||||||
GNB2L1 1084  ATAAAAAAA  1092
```

Figure B.15 Alignment data for GNB2L1 mRNA.

This insert sequence is homologous to nt 246-1092 of Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1 (GNB2L1) mRNA (NM_006098.4). This partial cDNA sequence is not fused in frame to the GAL4 DNA-BD and because of this it does not code a part or full of GNB2L1 protein.

Homo sapiens HCLS1 associated protein X-1 (HAX1), transcript variant 1, mRNA

```
Score = 1999 bits(1082), Expect  = 0.0, Identities = 1100/1108(99%)
Gaps = 3/1108(0%),Strand = Plus/Plus

Query  110   AGGGCTGCGCAGGTTTCCCCCGTCTGCGAATGGACCACTGGAGGGGTTCAAAGGTTCGCG  169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1   89    AGGGCTGCGCAGGTTTCCCCCGTCTGCGAATGGACCACTGGAGGGGTTCAAAGGTTCGCG  148

Query  170   TCCCAGTACGGGAATGAGCCTCTTTGATCTCTTCCGGGGCTTTTTCGGCTTTCCTGGACC  229
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    149     TCCCAGTACGGGAATGAGCCTCTTTGATCTCTTCCGGGGCTTTTTCGGCTTTCCTGGACC  208

Query   230     TCGGAGCCACAGAGATCCCTTTTTTTGGAGGGATGACTCGAGATGAAGATGATGATGAGGA  289
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    209     TCGGAGCCACAGAGATCCCTTTTTTTGGAGGGATGACTCGAGATGAAGATGATGATGAGGA  268

Query   290     AGAAGAAGAAGAAGGGGGCTCATGGGGCCGTGGGAACCCAAGGTTCCATAGTCCTCAGCA  349
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    269     AGAAGAAGAAGAAGGGGGCTCATGGGGCCGTGGGAACCCAAGGTTCCATAGTCCTCAGCA  328

Query   350     CCCCCCTGAGGAATTTGGCTTCGGCTTCAGCTTCAGCCCAGGAGGAGGGATACGTTTCCA  409
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    329     CCCCCCTGAGGAATTTGGCTTCGGCTTCAGCTTCAGCCCAGGAGGAGGGATACGTTTCCA  388

Query   410     CGATAACTTCGGCTTTGATGACCTAGTACGAGATTTCAATAGCATCTTCAGCGATATGGG  469
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    389     CGATAACTTCGGCTTTGATGACCTAGTACGAGATTTCAATAGCATCTTCAGCGATATGGG  448

Query   470     GGCCTGGACCTTGCCTTCCCATCCTCCTGAACTTCCAGGTCCTGAGTCAGAGACACCTGG  529
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    449     GGCCTGGACCTTGCCTTCCCATCCTCCTGAACTTCCAGGTCCTGAGTCAGAGACACCTGG  508

Query   530     TGAGAGACTACGGGAGGGACAGACACTTCGGGACTCAATGCTTAAGTATCCAGATAGTCA  589
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    509     TGAGAGACTACGGGAGGGACAGACACTTCGGGACTCAATGCTTAAGTATCCAGATAGTCA  568

Query   590     CCAGCCCAGGATCTTTgggggggTCTTGGAGAGTGATGCAAGAAGTGAATCCCCCCAACC  649
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    569     CCAGCCCAGGATCTTTGGGGGGGTCTTGGAGAGTGATGCAAGAAGTGAATCCCCCCAACC  628

Query   650     AGCACCAGACTGGGGCTCCCAGAGGCCATTTCATAGGTTTGATGATGTATGGCCTATGGA  709
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    629     AGCACCAGACTGGGGCTCCCAGAGGCCATTTCATAGGTTTGATGATGTATGGCCTATGGA  688

Query   710     CCCCCATCCTAGAACCAGAGAGGACAATGATCTTGATTCCCAGGTTTCCCAGGAGGGTCT  769
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    689     CCCCCATCCTAGAACCAGAGAGGACAATGATCTTGATTCCCAGGTTTCCCAGGAGGGTCT  748

Query   770     TGGCCCGGTTCTACAGCCCCAGCCCAAATCCTATTTCAAGAGCATCTCTGTGACCAAGAT  829
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    749     TGGCCCGGTTCTACAGCCCCAGCCCAAATCCTATTTCAAGAGCATCTCTGTGACCAAGAT  808

Query   830     CACTAAACCAGATGGGATAGTGGAGGAGCGCCGGACTGTGGTGGACAGTGAGGGCCGGAC  889
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    809     CACTAAACCAGATGGGATAGTGGAGGAGCGCCGGACTGTGGTGGACAGTGAGGGCCGGAC  868

Query   890     AGAGACTACAGTAACCCGACACGAAGCAGATAGCAGTCCTAGGGGTGATCCAGAATCACC  949
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    869     AGAGACTACAGTAACCCGACACGAAGCAGATAGCAGTCCTAGGGGTGATCCAGAATCACC  928

Query   950     AAGACCTCCAGCCCTGGATGATGCCTTTTCCATCCTGGACTTATTCCTGGGACGTTGGTT  1009
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    929     AAGACCTCCAGCCCTGGATGATGCCTTTTCCATCCTGGACTTATTCCTGGGACGTTGGTT  988

Query   1010    CCGGTCCCGGTAGCCTTGTTAACCCTCAGAGGCCTTCAAGTCCTTTCCACCTCTCACCCA  1069
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
HAX1    989     CCGGTCCCGGTAGCCTTGTTAACCCTCAGAGGCCTTCAAGTCCTTTCCACCTCTCACCCA  1048

Query   1070    TTGCCCTCCATTAATAAGCTTAGCTTCTCTTGCC-CCTCAGGGGCTTGGATATGTGGAAT  1128
                ||||||  |||||||||||||||||||||||||| ||||||||||||||||||||||||||
HAX1    1049    TTGCCCACCATTAATAAGCTTAGCTTCTCTTGCCACCTCAGGGGCTTGGATATGTGGAAT  1108

Query   1129    AATGAACTGGGGCCATGTCAGTTTGTCA-TCACC-AAACTGGCCAATAAAACCttttttt  1186
                | ||||||||||||||||||||||||| ||||| |||||| ||||||||||||||| |||
HAX1    1109    AGTGAACTGGGGCCATGTCAGTTTGTCACTCACCCAAACTGACCAATAAAACCTTTATTT  1168

Query   1187    ttgctaaaaaaaaaaaaaaaaaaaaa  1214
                 |||||||||||||||||||||||||
HAX1    1169    ATGCTAAAAAAAAAAAAAAAAAAAAA  1196
```

Figure B.16 Alignment data for HAX1 mRNA.

This insert sequence is homologous to nt 89-1196 of Homo sapiens HCLS1 associated protein X-1 (HAX1), transcript variant 1, mRNA (NM_006118.3). This partial cDNA sequence is not fused in frame to the GAL4 DNA-BD and because of this it does not code a part or full of HAX1 protein.

Homo sapiens chromosome 1 genomic scaffold, alternate assembly HuRef SCAF_1103279188134:57697-13195817

```
Score = 2111 bits(1143), Expect = 0.0, Identities =  1210/1242(97%)
Gaps = 7/1242(0%), Strand = Plus/Minus

Query  109      GGAGAGACAGGAAATCACACTTGGTATGACCTGTCTGTCACCTGGAGGTTTAGGTTTCCC  168
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3289031  GGAGAGACAGGAAATCACACTTGGTATGACCTGTCTGTCACCTGGAGGTTTAGGTTTCCC  3288972

Query  169      CAGCTTTACCAAGATGAGTGGGTATCCTCTATCCCTGCCAACCCCCACCTTGTGCTGACC  228
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288971  CAGCTTTACCAAGATGAGTGGGTATCCTCTATCCCTGCCAACCCCCACCTTGTGCTGACC  3288912

Query  229      CTTCCTTGGCCTGTCTTGGTTTTTTCCTTTCTTCCTTAAACAAAGCTACTCCAAATAGGA  288
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288911  CTTCCTTGGCCTGTCTTGGTTTTTTCCTTTCTTCCTTAAACAAAGCTACTCCAAATAGGA  3288852

Query  289      ACTAGGGTCCAAGGAAATAAAGACAGTATCCTAAAATggatttctcaatttcggcactat  348
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288851  ACTAGGGTCCAAGGAAATAAAGACAGTATCCTAAAATGGATTTCTCAATTTCGGCACTAT  3288792

Query  349      tgacattttTTGGGGGTGGGTGGGGAGGAAAGGGGCAGCCACTCACGCCACCACctctgc  408
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288791  TGACATTTTTTGGGGGTGGGTGGGGAGGAAAGGGGCAGCCACTCACGCCACCACCTCTGC  3288732

Query  409      ctgccactatcaacattttggactggataattctttgttgtgagaagctgtcctgtgcat  468
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288731  CTGCCACTATCAACATTTTGGACTGGATAATTCTTTGTTGTGAGAAGCTGTCCTGTGCAT  3288672

Query  469      tgtaggctgtttagcagcatttctgacctAGTAACTTCCCAGTTGTACAACTAAAAATGT  528
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288671  TGTAGGCTGTTTAGCAGCATTTCTGACCTAGTAACTTCCCAGTTGTACAACTAAAAATGT  3288612

Query  529      CTCCCTGGAAGGCAGAATTACCCTCTGTTGAGAACTATAGCCCTGGAAGAAATTGCCACC  588
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288611  CTCCCTGGAAGGCAGAATTACCCTCTGTTGAGAACTATAGCCCTGGAAGAAATTGCCACC  3288552

Query  589      TATACATGGGAAACTTAGAATTCAGGTGGAATAGGTTACTAGTTAAATATTTATGTAACA  648
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288551  TATACATGGGAAACTTAGAATTCAGGTGGAATAGGTTACTAGTTAAATATTTATGTAACA  3288492

Query  649      AGCAGCCTTCTGATAATCTGATTATGACTGTGTGTAGTGGCAACAACCACTTCTAAGTCA  708
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288491  AGCAGCCTTCTGATAATCTGATTATGACTGTGTGTAGTGGCAACAACCACTTCTAAGTCA  3288432

Query  709      GCTTTCCTTTAAGCCAGGCTCTGGATTCTTTATTTTTGCTGGCCTTAATCTGTGTCTGGC  768
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288431  GCTTTCCTTTAAGCCAGGCTCTGGATTCTTTATTTTTGCTGGCCTTAATCTGTGTCTGGC  3288372

Query  769      TTAACACCTCCTGGTCACTTTCCCTGGCTGGAATTTACGTACAGTTGTCAGTTGTTGATG  828
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288371  TTAACACCTCCTGGTCACTTTCCCTGGCTGGAATTTACGTACAGTTGTCAGTTGTTGATG  3288312

Query  829      CAGGTGTTTGGTTGATGATGCATTGCCAAAGAGAAGAGGAACTTCCTTTGTGGACATACT  888
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288311  CAGGTGTTTGGTTGATGATGCATTGCCAAAGAGAAGAGGAACTTCCTTTGTGGACATACT  3288252

Query  889      TCAGGGAGCTTCCTTGTTCTGTGCCTTTAAGGGGAGACTACAGTCTCTCAGTTGTTCTCC  948
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288251  TCAGGGAGCTTCCTTGTTCTGTGCCTTTAAGGGGAGACTACAGTCTCTCAGTTGTTCTCC  3288192

Query  949      ACAGCCAACCAAGAGTTTGGCTGCTGATATTTATAACCCTTATGTGACCACCCCCACCTG  1008
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288191  ACAGCCAACCAAGAGTTTGGCTGCTGATATTTATAACCCTTATGTGACCACCCCCACCTC  3288132
```

```
Query  1009     CCCACCCCAGAAAGAATTCCAATGCTCTGTAATAATCCTCTCCACCCCCACCATAGCAAC  1068
                ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288131  CCCACCCCACAAAGAATTCCAATGCTCTGTAATAATCCTCTCCACCCCCACCATAGCAAC  3288072

Query  1069     TGGATGTGGAAAAGGTGAAGGAGCCTTTAAGGCCTTTAAGTTTGCTTCATGTTGAACCTT  1128
                |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3288071  TGGATGTGGAGAAGGTGAAGGAGCCTTTAAGGCCTTTAAGTTTGCTTCATGTTGAACCTT  3288012

Query  1129     TTCCAGCTCCCTTTTAAAATATTGCCAGTTGGATTTTCAATTTCGGGGTCACAAGCAAAA  1188
                |||||||||||||||||||||||||||||| ||||| ||| || | ||||||||||||||
Sbjct  3288011  TTCCAGCTCCCTTTTAAAATATTGCCAGTTGTATTTTAAATATCAGTGTCACAAGCAAAA  3287952

Query  1189     TCCCAGGCTTGCAGCCACTGCTTCATCTTAAAGGAGCCTCATGCCCCATTGAAGGTAGTT  1248
                |||||||||||||||||||| |||||||||||||||||||||||| ||||||||||
Sbjct  3287951  TCCCAGGCTTGCAGCCACTGCATCATCTTAAAGGAGCCTCATGCCCCAGTGAAGGTAGTA  3287892

Query  1249     AATTTGA-TGGGAGGGTTTAAA-TCTTCNAGAAAGGATTGA-AGAATTTACCAAAA-CTG  1304
                || |||| ||| |||   |||| || || ||||||||||| ||| ||||| |||| |||
Sbjct  3287891  AAATTGAATGGAAGGTAGTAAAATCATCAAGAAAGGATTGATAGATTTTACAAAAATCTG  3287832

Query  1305     GCTGGGAATTTGA-AGGAAGGTGA-ATTTTAA-GCCACCAAC  1343
                 |||| ||||||| || ||||||| ||||||| || || |||
Sbjct  3287831  TCTGGAAATTTGATAGCAAGGTGAGATTTTAAAGCAACAAAC  3287790
```

Figure B.17 Alignment data for HuRef SCAF_1103279188134:57697-13195817.

This insert sequence is homologous to nt 3289031 – 3287790 of Homo sapiens chromosome 1 genomic scaffold, alternate assembly HuRef SCAF_1103279188134:57697-13195817 (NW_001838577.2). This sequence does not correspond to a transcript.

Homo sapiens chromosome X genomic scaffold, alternate assembly CHM1_1.0

```
Score = 1254 bits(679), Expect = 0.0, Identities = 679/679(100%)
Gaps = 0/679(0%), Strand = Plus/Plus

Query 109      TGAGTACACTGCTGGAAAGAGGGTAAACTGGGAGTTAGTGGATGGTCCCAATGCCCTGCC  168
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48761409 TGAGTACACTGCTGGAAAGAGGGTAAACTGGGAGTTAGTGGATGGTCCCAATGCCCTGCC  48761468

Query 169      TACAGCAGAGTGCCAACCAGCCCTGAGTGCAAAATTCAAGTTCAAtgtgtgtgcttgtgt  228
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48761469 TACAGCAGAGTGCCAACCAGCCCTGAGTGCAAAATTCAAGTTCAATGTGTGTGCTTGTGT  48761528

Query 229      gtggtgtgCTTTATGGACCCGCAAATACCATATTCATTATTGATGATAAGATCTTCACAG  288
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48761529 GTGGTGTGCTTTATGGACCCGCAAATACCATATTCATTATTGATGATAAGATCTTCACAG  48761588

Query 289      AATCCTGTAGCTACTAATGCATTGAGTTTTTAATCTCAGTACATCAGCCAGGAGGAGCCA  348
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48761589 AATCCTGTAGCTACTAATGCATTGAGTTTTTAATCTCAGTACATCAGCCAGGAGGAGCCA  48761648

Query 349      GATCACAGGGTAGTGATGTCTACTGGGATTATACTCATAACATCTACACAAAACAAGTTG  408
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48761649 GATCACAGGGTAGTGATGTCTACTGGGATTATACTCATAACATCTACACAAAACAAGTTG  48761708

Query 409      AGAAGGATCCACGTTTTCATTGTTTATCAGAATTGTATCTCATTTGGCTGAGCATTACTT  468
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48761709 AGAAGGATCCACGTTTTCATTGTTTATCAGAATTGTATCTCATTTGGCTGAGCATTACTT  48761768

Query 469      TTGTCAGAATGTGTTATCTGTAAACCATGTGTAGTGAAATTCTTCTGTAACTTTGGATTA  528
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48761769 TTGTCAGAATGTGTTATCTGTAAACCATGTGTAGTGAAATTCTTCTGTAACTTTGGATTA  48761828

Query 529      AAGGTATTTATGGTCTTTTTGTTTGTTTGATTTTTAAGTAAGTTATTTCTTTTGTAGACC  588
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48761829 AAGGTATTTATGGTCTTTTTGTTTGTTTGATTTTTAAGTAAGTTATTTCTTTTGTAGACC  48761888

Query 589      TGCTGATGGTATGGTTCCATCCTTCTGACCTCAGCATCCAATCTTTTTAAGGATTTTTGT  648
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48761889 TGCTGATGGTATGGTTCCATCCTTCTGACCTCAGCATCCAATCTTTTTAAGGATTTTTGT  48761948
```

```
Query 649      TTTCAATATTGTTATTTTAAATTGTGGTTGAAGCAATAGAAAATTGAAATATGGATTGTG  708
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48761949 TTTCAATATTGTTATTTTAAATTGTGGTTGAAGCAATAGAAAATTGAAATATGGATTGTG  48762008

Query 709      CATGACTGTGTCTTGAGTGTAAAAATATTGCAGTTTGAAACTTGGACCTAAAGTATTGCa  768
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 48762009 CATGACTGTGTCTTGAGTGTAAAAATATTGCAGTTTGAAACTTGGACCTAAAGTATTGCA  48762068

Query 769      aataaaaatgacaaacatc  787
               ||||||||||||||||||
Sbjct 48762069 AATAAAAATGACAAACATC  48762087
```

Figure B.18 Alignment data for Homo sapiens chromosome X genomic scaffold.

This insert sequence is homologous to nt 48761409 – 48762087 of Homo sapiens chromosome X genomic scaffold, alternate assembly CHM1_1.0 (NW_004078123.1). This sequence does not correspond to a transcript

Homo sapiens chromosome 14 genomic scaffold, alternate assembly CHM1_1.0 chr14

```
Score = 1042 bits(564), Expect = 0.0, Identities = 725/801(91%)
Gaps = 17/801(2%), Strand = Plus/Minus

Query 425      gtggtggctcacgcctggtaatctcagcgctttgggaggccgaggcaggtggatcactag  484
               |||||||||||||||| |||||| |||| |||||||||||| |||  ||  |||||  ||
Sbjct 39451876 GTGGTGGCTCACGCCT-GTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCAGATCATGAG  39451818

Query 485      gtcaggagttcgagaccagcctggccaacatggtgaaagcccatctctgctaaaaataac  544
               |||||||| | || |||| |||||| |||||||||||| ||||||||| |||||||||
Sbjct 39451817 GTCAGGAGATAGAAACCATCCTGGCTAACATGGTGAAACCCCATCTCTACTAAAAATACA  39451758

Query 545      aaggatcggccagatatgttggcatgtgcctgtaatcccagctgctcgggaggctgaggc  604
               ||  ||      |   || | ||  | ||||||| ||||| || |||||||||||||| |
Sbjct 39451757 AAAAATTAATTGGGCCTGGTAGC--GCGCCTGTAGTCCCACCTACTCGGGAGGCTGAGTC  39451700

Query 605      gggagagtcgcttgagcccgggaggcagaggttgcagtgagctaaggttgcaccactgca  664
                 ||||| |||||||| |||||||||  ||||||||  ||||| ||| |||| ||||||||
Sbjct 39451699 AGGAGAATCGCTTGAACCCGGGAGGTGGAGGTTGCCATGAGCCAAGATTGCGCCACTGCA  39451640

Query 665      ctccagcctggcgacaaagtgagactctgtctcaaaaacaaaaacaaaaacaaaaaacaa  724
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 39451639 CTCCAGCCTGGCGACAAAGTGAGACTCTGTCTCAAAAACAAAAACAAAAACAAAAAACAA  39451580

Query 725      aacaaaacaaaaatttttgttaagagggtagatctcacattaaatgttcttaacacacac  784
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 39451579 AACAAAACAAAAATTTTTGTTAAGAGGGTAGATCTCACATTAAATGTTCTTAACACACAC  39451520

Query 785      acacacacaaacacacacacacacacacacacacaca-aaa-gagacaaggagagcaGTC  842
               ||||||| | |||||||||||||||||||||||||| ||| ||||||||||||||||||
Sbjct 39451519 ACACACAAACACACACACACACACACACACACACACAAAAGAGACAAGGAGAGCAGTC  39451460

Query 843      CCCTGTAGGAAGAGACACTTTCCCTGGTCATGCTATCCAATAAATCACACCTGCTGTgga  902
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 39451459 CCCTGTAGGAAGAGACACTTTCCCTGGTCATGCTATCCAATAAATCACACCTGCTGTGGA  39451400

Query 903      gaaaggagaaaggaaaacaaggaaagagaaaaggatggaagctactagaaaaagaataga  962
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 39451399 GAAAGGAGAAAGGAAAACAAGGAAAGAGAAAAGGATGGAAGCTACTAGAAAAAGAATAGA  39451340

Query 963      aggCTGGGATGGAGAGAAAAGGGTTTCTCCCATACCCCTGATGTCTTTTTCTCTAATTAG  1022
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 39451339 AGGCTGGGATGGAGAGAAAAGGGTTTCTCCCATACCCCTGATGTCTTTTTCTCTAATTAG  39451280

Query 1023     AAA-GACAGTGCTTTTCTTTCCAACCTCTTTAGTTTTTTCTTTCCCATAAGCTGTTTCTT  1081
               ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 39451279 AAAAGACAGTGCTTTTCTTTCCAACCTCTTTAGTTTTTTCTTTCCCATAAGCTGTTTCTT  39451220

Query 1082     TCAGTAGCTGCTAGCTTTAATTATATGCAATAATTGGAAGTTAATTAAGTATCCCAAGGG  1141
               |||||||||||||||| | |||||||||||| ||||| |||||||| |||||||| |||
```

```
Sbjct 39451219 TCAGTAGCTGCTAGCTTAATTTATATGCAATATTTGGA-GTTAATTA-GTATCCCA-GGG  39451163

Query 1142     TAACCGAATTCCCTTACCGAAGCCCTTTAGCTGGGGCCTTCCGCTGGAAAAAATGCCCTC  1201
               |||| || |||||||| ||| ||| || ||||| | ||||| |||| | ||| |||| ||
Sbjct 39451162 TAACAGA-TTCCCTTAGCGATGCC-TT-AGCTGTG-CCTTCAGCTG-ATAAACTGCC-TC  39451109

Query 1202     AAATACAAAttttttccaaaaa  1222
               |||||||| |||| ||| |||
Sbjct 39451108 AAATACAA-TTTT-CCAGAAA  39451090
```

Figure B.19 Alignment data for Homo sapiens chromosome 14 genomic scaffold.

This insert sequence is homologous to nt 39451876 – 39451090 of Homo sapiens chromosome 14 genomic scaffold, alternate assembly CHM1_1.0 chr14 (NW_004078079.1). This sequence does not correspond to a transcript.

Homo sapiens chromosome 2 genomic scaffold, alternate assembly CHM1_1.0

```
Score = 1495 bits(809), Expect = 0.0, Identities = 814/816(99%)
Gaps = 1/816(0%), Strand = Plus/Minus

Query  107     GGACCAATCAAAACCGTTACCTCTGATAACATTATATTGTGCCCAGTATAAAATAGATAT  166
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162824  GGACCAATCAAAACCGTTACCTCTGATAACATTATATTGTGCCCAGTATAAAATAGATAT  162765

Query  167     AATAATACCTCAAACTTAAATCCGGGCATTGTCATTGAATATCTTAAGAATATGCAACAA  226
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162764  AATAATACCTCAAACTTAAATCCGGGCATTGTCATTGAATATCTTAAGAATATGCAACAA  162705

Query  227     AGGTGCTTTTAAAAATACAAGCtagtgattgtaccaaatttgtaaatcacataggatagt  286
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162704  AGGTGCTTTTAAAAATACAAGCTAGTGATTGTACCAAATTTGTAAATCACATAGGATAGT  162645

Query  287     gggtcattttaagaatattagttatttcaatctataaacgtggatgtctttccttttttg  346
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162644  GGGTCATTTTAAGAATATTAGTTATTTCAATCTATAAACGTGGATGTCTTTCCTTTTTTG  162585

Query  347     tgttttctttaatttctttcattaatatttgtcatttttgttgtcgaaatcttttacttc  406
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162584  TGTTTTCTTTAATTTCTTTCATTAATATTTGTCATTTTTGTTGTCGAAATCTTTTACTTC  162525

Query  407     cttggttaaatttatttctaagtacatttttgtagctattgtaaaaggaattgctttctt  466
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162524  CTTGGTTAAATTTATTTCTAAGTACATTTTTGTAGCTATTGTAAAAGGAATTGCTTTCTT  162465

Query  467     aatttcttgtttcagctagtttactatcaatatatagaaatgctactgatttttgtatgt  526
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162464  AATTTCTTGTTTCAGCTAGTTTACTATCAATATATAGAAATGCTACTGATTTTTGTATGT  162405

Query  527     tgatttatatcctgcaactttattaatttcatgtatcaccctaagaagctttggtagag  586
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162404  TGATTTATATCCTGCAACTTTATTAATTTCATGTATCACCCTAAGAAGCTTTTGGTAGAG  162345

Query  587     tcttatttttttccgtgtataagatcacattgtctttaaacagggacaatttgactgtct  646
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162344  TCTTATTTTTTTCCGTGTATAAGATCACATTGTCTTTAAACAGGGACAATTTGACTGTCT  162285

Query  647     cctttccaattcagatgtcctttatttctttctctcacctaattgtcctggctaagactt  706
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162284  CCTTTCCAATTCAGATGTCCTTTATTTCTTTCTCTCACCTAATTGTCCTGGCTAAGACTT  162225

Query  707     tcactatgtgaaatatgattggtgagaataggcatccttttcttgttccagtaaaatcct  766
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct  162224  TCACTATGTGAAATATGATTGGTGAGAATAGGCATCCTTTTCTTGTTCCAGTAAAATCTT  162165

Query  767     tttcttgttcacagtaaaatctttcacctttttccacactcagtatgatcttagttgtaga  826
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162164  TTTCTTGTTCACAGTAAAATCTTTCACCTTTTCCACACTCAGTATGATCTTAGTTGTAGA  162105

Query  827     tttgtcctttatgtccttctgttttaaggcatatattttctatactaaattgttgagagg  886
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  162104  TTTGTCCTTTATGTCCTTCTGTTTTAAGGCATATATTTTCTATACTAAATTGTTGAGAGG  162045
```

```
Query  887     ttttttgtcatgtaagaatatttaattttgccgaaa  922
               ||||||||||||||||||||||||||||||||| |||
Sbjct  162044  TTTTTTGTCATGTAAGAATATTTAATTTTGCC-AAA  162010
```

Figure B.20 Alignment data for Homo sapiens chromosome 2 genomic scaffold.

This insert sequence is homologous to nt 162824 – 162010 of Homo sapiens chromosome 2 genomic scaffold, alternate assembly CHM1_1.0 (NW_004078007.1). This sequence does not correspond to a transcript.

CURRICULUM VITAE

Name Surname: Ahmet Erdal YILMAZ

Place and Date of Birth: Istanbul, 1984

E-Mail: erdalyilmaz@itu.edu.tr

B.Sc.: Ege University, Bioengineering Department

M.Sc.: Istanbul Technical University
Molecular Biology Genetics and Biotechnology Programme